Atomic Energy Control Board
P.O. Box 1046
Ottawa, Canada
K1P 5S9

Commission de contrôle de l'énergie atomique
C.P. 1046
Ottawa, Canada
K1P 5S9

Canada

INFO-0172

UNCERTAINTY ANALYSIS AND PROBABILISTIC MODELLING

by

SENES Consultants Limited

A report prepared for the
Atomic Energy Control Board
Ottawa, Canada

May, 1985

Research report

UNCERTAINTY ANALYSES AND PROBABILISTIC MODELLING

ABSTRACT

Many factors affect the accuracy and precision of probabilistic assessment. This report discusses sources of uncertainty and ways of addressing them. Techniques for propagating uncertainties in model input parameters through to model prediction are discussed as are various techniques for examining how sensitive and uncertain model predictions are to one or more input parameters. Various statements of confidence which can be made concerning the prediction of a probabilistic assessment are discussed as are several matters of potential regulatory interest.

RÉSUMÉ

De nombreux facteurs influent sur l'exactitude et la précision de l'évaluation probabiliste. Le présent rapport traite des sources d'imprécision et des moyens d'y remédier. Les techniques de propagation des imprécisions dans les paramètres de données de modèles jusque dans les prévisions de modèles sont discutées, ainsi que diverses techniques pour examiner à quel point les prévisions de modèles sont sensibles et imprécises par rapport à un ou plusieurs paramètres de données. Divers énoncés qui peuvent être faits sur la fiabilité d'une évaluation probabiliste sont traités, comme le sont plusieurs questions susceptibles de présenter un intérêt en matière de réglementation.

DISCLAIMER

The Atomic Energy Control Board is not responsible for the accuracy of the statements made or opinions expressed in this publication and neither the Board nor the authors assume liability with respect to any damage or loss incurred as a result of the use made of the information contained in this publication.

ACKNOWLEDGEMENTS

This report was prepared under the direction of Mr. B. E. Halbert of SENES Consultants Limited. Other SENES staff who made significant technical contributions to the report include Mr. D.I. Beals, Dr. D.B. Chambers, Dr. L.M. Lowe and Mr. M.L. Murray. The expertise provided by a number of specialist advisors to SENES is gratefully acknowledged. These include Dr. F.O. Hoffman and Dr. R.H. Gardner on techniques of sensitivity and uncertainty analysis, and Dr. S. Kaplan on the method of discrete probability distributions. We are particularly grateful to Professor P. Reilly and Mr. T. Duever for their contributions on random sampling methods and for Professor Reilly's many useful suggestions concerning the draft manuscript. Finally, we also wish to acknowledge the assistance of Ms. S. Dey and Mr. D. McCaffrey of SENES who actually assembled the report.

Table of Contents

Page

ABSTRACT i

ACKNOWLEDGEMENTS ii

1.0 INTRODUCTION
- 1.1 Study Objectives 1-1
- 1.2 Some General Considerations Related to Predictive Modelling 1-2
- 1.3 Study Approach and Report Structure 1-6

2.0 UNCERTAINTY IN ASSESSMENT MODELLING
- 2.1 Sources of Uncertainty 2-1
- 2.2 Model Uncertainties 2-2
  - 2.2.1 Imprecisions in Model Formulation 2-2
  - 2.2.2 Approximations in Model Implementation 2-4
- 2.3 Parameter Uncertainties 2-6
  - 2.3.1 Measurement Uncertainty in Parameter Values 2-6
  - 2.3.2 Inherent Variability in Parameter Values 2-7
  - 2.3.3 Imprecision in Parameter Estimation 2-9
  - 2.3.4 Correlations Between Parameters 2-10
- 2.4 Some Other Sources of Uncertainty 2-12
- 2.5 Further Observations Regarding Uncertainty 2-13

3.0 CHARACTERIZING AND PROPAGATING UNCERTAINTY
- 3.1 Scenario Analysis 3-1
- 3.2 Model Selection 3-2
- 3.3 Analytic Methods 3-4

Table of Contents (continued)

Page

3.4 Simple Random Sampling and Alternatives 3-10
3.4.1 Introduction 3-10
3.4.2 Simple Random Sampling (RS) 3-11
3.4.3 Stratified Random Sample (SRS) 3-13
3.4.4 Quota Sampling 3-14
3.4.5 Systematic Sampling (SYS) 3-15
3.4.6 Latin Hypercube Sampling (LHS) 3-16
3.4.7 Comparison of Sampling Methods with Respect to Monte Carlo Calculations 3-18
3.4.8 Suggested Sampling Methods 3-21
3.4.9 Some Examples 3-21

4.0 SENSITIVITY AND UNCERTAINTY ANALYSIS TECHNIQUES

4.1 General Discussion of Techniques 4-1
4.1.1 Tomovic Sensitivity and the Slope of the Line 4-2
4.1.2 Regression Analysis Techniques 4-4
4.1.3 Sensitivity Index 4-10
4.1.4 Graphical Techniques 4-11
4.1.5 Uncertainty Index 4-12
4.1.6 Other Techniques 4-14
4.2 Some Examples Used to Illustrate Methods of Sensitivity and Uncertainty Analysis 4-14
4.2.1 Model Y1 4-15
4.2.2 Model Y2 4-16
4.2.3 Model Y3 4-17
4.2.4 Model Y5 4-19
4.3 Discussion 4-20

Table of Contents (continued)

Page

5.0 CONSIDERATIONS IN QUANTIFYING CONFIDENCE IN PROBABILISTIC MODELLING RESULTS

5.1 Some Concepts of Probabilistic Modelling 5-1
5.2 Interval Estimates 5-4
5.3 Confidence Bands 5-8

6.0 SOME ASPECTS OF REGULATORY CONCERN

6.1 Introduction 6-1
6.2 Interpretation of the Output of a Probabilistic Assessment 6-1
6.3 The Concept of the Critical Group 6-6

7.0 CONCLUDING REMARKS

REFERENCES

APPENDIX A - NOTES ON SAMPLING FROM RANDOM DISTRIBUTIONS

APPENDIX B - THE METHOD OF DISCRETE PROBABILITY DISTRIBUTIONS

List of Figures

		Follows Page
3.1	Components of Error in Model Predictions As a Function of Model Complexity	3-3
3.2	Histogram of Y1=A+B+C+D+E with A Uniformly Distributed From 1 to 3 and B=C=D=E=2.0	3-6
3.3	Histogram of Y1=A+B+C+D+E with A and B Uniformly Distributed from 1 to 3 and C=D=E=2.0	3-6
3.4	Histogram of Y1=A+B+C+D+E with A, B and C Uniformly Distributed from 1 to 3 and D=E=2.0	3-6
3.5	Histogram of Y1=A+B+C+D+E with A, B, C, and D Uniformly Distributed from 1 to 3 and E=2.0	3-6
3.6	Histogram of Y1=A+B+C+D+E with All Inputs Uniformly Distributed from 1 to 3	3-6
3.7	Simple Random Sampling (RS)	3-11
3.8	Simple Random Sampling Cumulative Probability Function for A	3-11
3.9	Stratified Random Sampling (SRS)	3-13
3.10	Latin Hypercube Sampling (LHS)	3-17
3.11a	Histogram of A (Triangular Distribution) Obtained Using RS	3-22

List of Figures (Continued)

		Follows Page
3.11b	Histogram of A (Triangular Distribution) Obtained Using LHS	3-22
3.12a	Histogram of B (Uniform Distribution) Obtained Using RS	3-22
3.12b	Histogram of B (Uniform Distribution) Obtained Using LHS	3-22
3.13a	Histogram of C (Loguniform Distribution) Obtained Using RS	3-22
3.13b	Histogram of C (Loguniform Distribution) Obtained Using LHS	3-22
3.14a	Histogram of D (Lognormal Distribution) Obtained Using RS	3-22
3.14b	Histogram of D (Lognormal Distribution) Obtained Using LHS	3-22
3.15a	Histogram of E (Normal Distribution) Obtained Using RS	3-22
3.15b	Histogram of E (Normal Distribution) Obtained Using LHS	3-22
3.16a	Histogram of Y1 Obtained Using RS	3-22
3.16b	Histogram of Y1 Obtained Using LHS	3-22

List of Figures (Continued)

		Follows Page
3.17a	Histogram of Y2 Obtained Using RS	3-22
3.17b	Histogram of Y2 Obtained Using LHS	3-22
3.18a	Histogram of Y3 Obtained Using RS	3-22
3.18b	Histogram of Y3 Obtained Using LHS	3-22
3.19a	Histogram of Y4 Obtained Using RS	3-22
3.19b	Histogram of Y4 Obtained Using LHS	3-22
3.20a	Histogram of Y5 Obtained Using RS	3-22
3.20b	Histogram of Y5 Obtained Using LHS	3-22
4.1	Sensitivity of Output to Single Input Parameter	4-1
4.2	Illustration of Sensitivity and Uncertainty of Output to Two Input Parameters	4-3
5.1	Histogram of Y5 with Case V Inputs and 50 Iterations	5-2
5.2	Histogram of Y5 with Case V Inputs and 100 Iterations	5-2
5.3	Histogram of Y5 with Case V Inputs and 200 Iterations	5-2

List of Figures (Continued)

		Follows Page
5.4	Histogram of Y5 with Case V Inputs and 500 Iterations	5-2
5.5	Cumulative Relative Frequency (500 Trials of Model Y5 with Mixed Parameters)	5-3
5.6	Complementary Cumulative Relative Frequency (500 Trials of Model Y5 with Mixed Parameters)	5-3
6.1	Typical Output from a Probabilistic Model	6-1
A.1a	The Exponential Probability Density Function f(x)	A-2
A.2b	The Cumulative Distribution Function F(x)	A-2
B.1	Cumulative Histograms of Model Y5 Calculated Using RS and DPD	B-2

List of Tables

		Follows Page
2.1	Uncertainty in Assessment Models	2-1
3.1	Propagation of Error Formulas for Some Simple Functions	3-5
3.2	Example Models Used in Comparison of Latin Hypercube and Simple Random Sampling Methods	3-22
3.3	Input Parameter Distributions Used in Example Models	3-22
3.4	Comparison of Means and Variances of Example Models	3-22
4.1	Example Models Used to Illustrate Techniques of Sensitivity/Uncertainty Analysis	4-15
4.2	Input Parameter Distributions Used in Example Models	4-15
4.3	Sensitivity/Uncertainty Indices for Model Y1 for Five Cases	4-15
4.4	Sensitivity/Uncertainty Indices for Model Y2 for Five Cases	4-16
4.5	Partial Derivatives and Regression Information for Model Y3 for Five Cases	4-17
4.6	Sensitivity/Uncertainty Indices for Model Y3 for Five Cases	4-17

List of Tables (Continued)

Follows Page

4.7 Partial Derivatives and Regression Information for Model Y5 for Five Cases 4-19

4.8 Sensitivity/Uncertainty Indices for Model Y5 for Five Cases 4-19

B.1 Comparison of Means and Variances of Example Models Calculated Using DPD and RS B-2

1.0 INTRODUCTION

1.1 Study Objectives

The acceptability of facilities and methods proposed for the disposal of radioactive waste will likely be assessed, in part at least, on the basis of results of predictive modelling. Traditionally, such predictive models have been used, after selection of a single value for each of the input parameters, to arrive at a single estimate of some environmental consequence such as dose to an individual in the critical group. This method of calculation is often classified as a deterministic model evaluation.

By their very nature however, models are only imperfect reflections of real systems. Moreover, many of the parameter values used as inputs in model evaluations are either incompletely known or inherently variable. As a result, over the past decade, considerable attention has been given to incorporating the uncertainty (and variability) associated with individual parameters directly into the model predictions. This is often done by sampling from specified distributions of input parameters to arrive at an estimated distribution of the environmental consequence of interest. The mathematical models employed in such analyses are usually no different than those formerly used. It is simply that they are being solved in a probabilistic rather than in a deterministic manner.

The interpretation of a probabilistic analysis is more complex than that of a deterministic analysis. Decision makers must also consider the uncertainty associated with estimates of environmental consequence as well as the magnitudes of the estimates themselves. Uncertainties in the probabilistic results are due in part to many of the same factors which contribute to uncertainty in the deterministic results (e.g. model selection or parameter value estimation). The difference is that the uncertainties in the parameter values are explicitly considered in a probabilistic analysis. In addition, the mechanical, calculational aspects of carrying out a probabilistic assessment affect the nature of the output values and need to be considered.

Because of these difficulties of interpretation and recognizing that decisions concerning the performance of radioactive waste disposal facilities will likely have to be made on the basis of predictive modelling, some of which may be probabilistic in nature, the Atomic Energy Control Board (AECB) commissioned the present study in November 1984 with the objective:

> "...to discuss and clarify the factors which affect the accuracy and precision of quantities calculated by probabilistic predictive models, and to detail methods for quantifying uncertainty, with a view to quantifying the confidence which can be assigned to model output predictions."

This report presents the results of the study undertaken for the AECB by SENES Consultants Limited.

1.2 Some General Considerations Related to Predictive Modelling

The phrases "predictive model" and "mathematical model" can be used in many different ways. In this report, we use these phrases to mean an explicit symbolic formulation relating one or more independent variables to the dependent variable, along with a specification of all inputs (i.e. initial conditions and parameters) necessary to evaluate the model. For example, the equation

$$\frac{dY}{dt} = -K\ Y(t) \qquad (1\text{-}1)$$

is a mathematical model of the physical situation existing when a quantity of radioactive material undergoes decay. In this case, the amount of radioactive material, Y(t), remaining at any time t is determined by the solution to equation (1-1) in accordance with the specification of input data which includes the initial condition, namely, the initial amount of the radioactive material, Y(t=0), and the parameter K, the radioactive decay constant.

The term "model" is also used in a second sense in this report, namely, in the sense of a regression model where a mathematical regression is carried out on a set of data to determine the relation between the dependent variable Y and the independent (regression) variables X_i. For example, in the case of a linear regression Y on the set of input variables $<X_i>$

$$Y = a_o + b_1X_1 + b_2X_2 + \ldots \qquad (1\text{-}2)$$

the $<X_i>$ are the independent variables, Y is the dependent variable, and a_o and the $<b_i>$ are the parameters determined through the regression analysis. In many instances throughout this report, the techniques of regression analysis are applied to the predicted outputs generated using mathematical models that are intended to reflect physical concepts. The regression equation (1-2) however, is purely a mathematical construct, and does not have a fundamental "physical meaning", as say equation (1-1) does.

The independent variables in a regression analysis are most often the same "parameters" (which could be considered as possibly adjustable constants) used in the "mathematical model".

Regression analysis is used to help analyze the results. The meaning of the terms "model variable" and "model parameter" should hopefully be clear from the context of the particular discussion.

The ultimate objective of a modelling exercise is to produce a set of predictions which accurately reflect the true behaviour of a real system. However, predictions made with mathematical models will not always agree with observational evidence (if any exists). Setting aside any computational errors, the differences may be due either to error or to discrepancies between the mathematical model and the actual situation which the model purports to represent. As discussed later in Chapter 2, several types of uncertainty contribute to the overall performance of the model including uncertainty present in the input distributions selected to characterize the values of the model parameters. In this regard, model sensitivity and uncertainty analysis techniques can be used to provide estimates of the relative importance of each model parameter.

At this point, it is worth noting the distinction between sensitivity and uncertainty analysis as used in this report. "Sensitivity analysis" refers to the sensitivity of the magnitude of the model output to changes in the magnitudes of the input parameters. "Uncertainty analysis" refers to how much the uncertainty (or spread) associated with each parameter contributes to the total uncertainty (or spread) of the model result.

The inferences which can be drawn from sensitivity and uncertainty analysis are directly dependent upon the numerous assumptions and compromises made necessary because of data inadequacies, imperfections in the mathematical formulation of the model, and incomplete understanding of the detailed function and structure of real systems. The set of conditions defined by these assumptions and compromises forms the "framework" or the "context" within which any analysis of sensitivities or uncertainties must be conducted. Thus, changes in any modelling component (e.g. changes to the model structure, new data, or revisions in the objectives of the assessment) will form a new "context" for analyzing sensitivities and uncertainties.

Such considerations are important to both the modeller and the regulator, particularly in trying to answer questions of the type "how good is good enough" when analysing the results of modelling exercises. As illustrations of this view, consider that:

1. The process of model development and analysis is ongoing and terminates only when someone decides that current predictions are "good enough" for some purpose or when further time and/or funds are lacking.

2. Meaningful results are difficult to obtain from any model evaluation technique without first explicitly considering the "context" of model application (i.e. what exactly it is we wish to model and what conclusions we would like to be able to draw from the predictive modelling?) .

3. The results of uncertainty and sensitivity analysis should play a key role in identifying the most effective areas where further model development and data acquisition can be directed to enhance the accuracy of the model's predictions.

4. The regulatory interpretation of results from a probabilistic analysis is more complex than for a deterministic analysis. In a deterministic analysis, the regulator is faced with a single estimate of consequence which can be compared to a single regulatory limit. In

probabilistic analysis, the regulator is presented with a distribution of consequence values. These aspects are discussed at more length in Chapter 6.0.

1.3 Study Approach and Report Structure

Mathematical models have been and will continue to be essential tools for assessing the environmental consequences of radioactive waste disposal. Uncertainty, in many forms, is inherent in such assessments and must be recognized and dealt with. Recognition of the various potential sources and types of uncertainty is thus fundamental to any discussion of uncertainty in predictive modelling.

A qualitative discussion of many of the sources of uncertainty is provided in Chapter 2. Subsequent sections of the report deal mainly with uncertainty in model input parameters. The assessment models are assumed to be true representations of the real systems they are intended to reflect (i.e. they have been validated in some sense).

Chapter 3 presents a number of techniques which have been used for propagating uncertainty within predictive models. Specific numeric examples are provided to illustrate in a quantitative fashion various aspects of the techniques.

Chapter 4 discusses methods for looking at the results of model predictions with a view to determining how sensitive and uncertain the model predictions are to the (changes in) various model parameters.

Chapter 5 addresses questions related to how confident, in a statistical sense, one can feel about probabilistic model predictions.

Chapter 6 addresses several questions which have particular relevance to how a regulator might interpret the results of a probabilistic assessment.

Chapter 7 provides a few concluding remarks concerning the major observations obtained from the study. Chapter 7 is followed by references and Appendices A and B which discuss, respectively, some methods for sampling from random distributions, and the method of discrete probability distributions.

The intent of this report is to provide a general overview of the sources of uncertainty and practical ways to deal with them. The report is not intended to be a treatise which sets out either the theoretical basis for particular techniques or to describe in detail how to apply the various techniques discussed herein. The reader interested in such information is referred to the numerous publications cited in the report.

2.0 UNCERTAINTY IN ASSESSMENT MODELLING

2.1 Sources of Uncertainty

As noted earlier, environmental assessment models are, by their very nature, imperfect reflections of real systems. Hence there are inherent uncertainties in the model results which are attributable to a number of sources and factors. The primary sources of uncertainty are associated either with errors in the mathematical model construction or uncertainties in the model parameters. Table 2.1 lists sources of uncertainty which can arise in assessment modelling. Some obvious potential sources of error such as computer malfunction are not relevant to the present discussion and hence are not shown in Table 2.1.

The distinction betweeen a parameter and a model can be fuzzy, depending in part on the intended application of the model. For example, consider the parameter "bioaccumulation factor" (BF) which is commonly used to represent the net transfer of radioactivity from water to fish flesh. Scrutiny of the "parameter" BF reveals that it is a surrogate for a complex array of processes whereby radioactivity is transferred via a variety of pathways from the water to the fish flesh. From this perspective, the "parameter" BF can be considered as an aggregate "model".

As noted above, the primary sources of uncertainty are model uncertainty and parameter uncertainty. Model uncertainties (discussed in Section 2.2) arise from basic incommensurability between the real and the mathematical worlds. Parameter uncertainties (discussed in Section 2.3) arise, and manifest themselves, in a variety of ways. First, there is the difficulty of measuring parameter values accurately. Aside from the purely mechanical problems of measurement accuracy, uncertainty is also

Table 2.1

Uncertainty in Assessment Models

Source	Discussed in Section
- scenario uncertainty	3.1
- model selection and formulation	2.2.1, 3.2
- model implementation	
- approximations	2.2.2
- round-off errors	2.2.2
- number of trials	5.1, 5.2, 5.3
- random number generation	2.2.2
- software errors	2.2.2
- parameter uncertainties	
- measurement uncertainty	2.3.1
- variability in time and space	2.3.2, 6.2
- use of inappropriate parameter values	2.3.3
- parameter correlation	2.3.4
- variations among expert groups	2.3.3, 2.4

introduced by using parameters measured or defined for one particular application in some different (and perhaps inappropriate) analysis. Additionally, the analysis is complicated by the fact that correlations may exist among various of the physical parameters; the form or magnitude of such correlations is often uncertain. Finally, some parameters are stochastic in nature and thus are inherently variable; wind speed, for example, is a naturally variable parameter. These and several others sources of uncertainty in mathematical modelling, are discussed below.

While sources of uncertainty in model predictions are well documented in the literature, less attention has been given to identifying methods of quantifying the magnitudes of these uncertainties. Where techniques exist to quantify sources of uncertainty, they are mentioned in the appropriate sections.

2.2 Model Uncertainties

2.2.1 Imprecisions in Model Formulation

Formulation of an appropriate mathematical model is likely the most critical and difficult step in any modelling exercise. Hoffman and Miller (1983) claim that "incorrect model formulation is the primary, underlying source of uncertainty," since the model structure will influence all modelling results, regardless of the care taken in selecting input parameters. Given the complex nature of the physics and chemistry of contaminant source terms, environmental transport mechanisms, food chain bioaccumulation processes and human dosimetry, a variety of scientific models can be formulated to describe any given system.

Thus, the modeller must recognize which features and processes form the fundamental basis for the system of interest and incorporate appropriate mathematical descriptions of these in the model. Any intentional or inadvertent departures of model formulations (including implicit and explicit assumptions) from real world conditions will result in systematic over- or under-prediction of real world events, a phenomenon known as model bias. Intentional sources of model bias are usually related to the application of conservative assumptions. Unintentional sources of model bias, on the other hand, "may be related to unwarranted assumptions of steady-state conditions or first-order kinetics, unrecognized relationships between model parameters or inadvertent exclusion of significant exposure pathways or important physical processes" (Hoffman and Miller 1983).

It should be realized that a model formulated correctly for one set of conditions may not perform as well when extrapolations are made to new situations. It is also worth noting that, given a number of alternative functional model forms, it may be possible in many instances to reformulate them as "a single form with an extra parameter that can make the new model form equivalent to each of the original forms according to the value chosen" (Henrion and Morgan 1984). For example, in toxicology it is possible to define a dose-response function with a threshold parameter and a dose-exponent parameter. If the threshold parameter is zero, then the dose-response function will simulate a non-threshold model i.e. there could be a response (such as an observed health effect) right down to zero dose. If the threshold parameter is non-zero, the function will simulate a model with a threshold. Likewise, the dose-exponent parameter will define whether the dose-response function is linear or non-linear with dose. Thus, uncertainty about the model form may be converted into uncertainty about parameter values.

Uncertainties in model formulation can be quantified by comparing model predictions with field observations within the intended range of model application. This comparative method, known as model validation, is helpful for many models and model uses but may be difficult or impossible to apply to other model applications (for example, for predictions far into the future). Atmospheric dispersion models, such as the Gaussian plume model, have been subjected to extensive validation but large, complex radiological assessment models do not as readily lend themselves to validation. The models of interest to this study would typically deal with low environmental concentrations of radionuclides and consequent doses to humans over long time periods (perhaps thousands of years into the future) and hence would be impossible to validate completely. It may only be practical to validate smaller models (such as dispersion models) which may be more tractable.

One widely used approach to model validation involves comparing model calculations to laboratory and field data. It should be noted, however, that this method commonly requires splitting the experimental data into two parts, one for estimating unknown model parameters and the other for model validation. In order to help overcome the difficulties of evaluating model performance over long time periods (a feature not usually possible in experimental work), reference can be made to natural analogues. For example, studies of groundwater in close proximity to uranium ore bodies can provide insight as to how radionuclides migrate away from the ore body.

2.2.2 Approximations in Model Implementation

Once the functional form of a model has been selected, approximations must generally be made (and uncertainty introduced) to allow for computationally feasible implementation. This generally involves discretizing continuously variable systems by introducing finite spatial or temporal grids (Henrion and Morgan 1984). Prime examples of this are the consideration of wind speed and turbulence (i.e. dispersive capacity of the environment) in terms of stability classes, and modelling wind direction in terms of 16 points of a compass (Fox 1984).

The finite number of runs used in a Monte Carlo simulation also introduces uncertainty, since it is not likely that all possible combinations of input parameter values will ever be incorporated. This source of uncertainty is discussed in Chapter 5. Implementation uncertainty of a different nature is introduced during the computational solution of the mathematical equations. According to Hendrickson (1984), "...errors introduced by rounding and truncation arise primarily from register lengths in computer hardware, and are entirely dependent on the mechanical features of the equipment used for the model computation. They may or may not be important to sensitivity analysis since most computers provide floating point calculations. It is not unreasonable to make the point that, for most simulations, the input data are far less accurate than the capacity for accuracy provided by most modern computers. However, round-off error is an extremely important subject and has been treated widely in the literature as it relates to and impacts on the accuracy of repeated calculations, convergence criteria, and decisions based on numerical differences."

The effects of many of the implementation approximations are best examined by changing the level of detail in the implementation. That is, the effects of altering spatial grid size, magnitude of time steps, number of Monte Carlo trials, etc., should all be examined. Henrion and Morgan (1984) point out that "With many computer models, this requires rather fundamental re-programming, and so it is relatively rare that modellers carry out such examination in a systematic way. This is unfortunate since a comparison of uncertainty introduced by model implementation and simulation with other sources of uncertainty can help analysts select an appropriate level of model detail. "Appropriate" here means neither so complex that the computational burden is overwhelming, nor so simple that approximation uncertainty dominates all other sources."

Widespread use is made of computer-generated random numbers. Since random number generation can be affected by both hardware and software characteristics, it is important that the random number generator used in a simulation be implemented appropriately for the computer environment in which it is used. The interested reader is referred to the literature for discussions of portability considerations (e.g. Gentle 1981, Morse and Roberts 1983).

Software errors are an important consideration in helping to ensure that the results of an assessment modelling exercise can be interpreted in a meaningful way. Software quality assurance implies constantly asking and answering the question "Am I building the product right?". This implies not only that the computer code is (reasonably) error free and performs according to some preconceived specifications but that the specifications themselves were properly set out in the first place - in other words, does the software actually do what it was originally intended to do? The reader interested in further discussion of this important topic is referred to the extensive literature (e.g. Boehm 1984, Myers 1979).

2.3 Parameter Uncertainties

2.3.1 Measurement Uncertainty in Parameter Values

As is well known in any scientific endeavor, it is not possible to measure any physical quantity with complete certainty.

Ignoring for the moment, the random variability of many physical quantities, the uncertainty associated with a measured physical quantity will depend greatly on both the nature of the physical quantity itself and the state-of-the-art of instrumentation technology for measuring that particular quantity. The accuracy with which many quantities can be measured is constantly improving (Hendrickson 1984). The random error in measurement can be quantified using standard statistics such as standard deviation, or percentiles of a frequency distribution. Such uncertainty is common to all branches of science, is well understood and the tools for estimating it are well developed (Henrion and Morgan 1984).

Cranwell and Helton (1981) point out that in cases where relatively sophisticated measurements are required, the measurement techniques used may be inappropriate or misapplied. This is especially true where laboratory analysis of environmental samples is required. If the samples are stored inappropriately or analyzed under conditions significantly different from those found in nature, the results of any analysis may be in error.

2.3.2 Inherent Variability in Parameter Values

As with uncertainty associated with parameter measurement, it is commonly accepted that uncertainty or variability is inherent in natural systems. Inherent uncertainty arises from considering time or spatial averages of a sequence of instantaneous measurements on stochastic parameters. Parameters describing atmospheric motion, flow conditions in surface water drainage systems or numerous soil parameters are all common examples of inherently random or stochastic quantities. As noted by Fox (1984) and Hofer and Krzykacz (1984), this type of uncertainty arises from the fact that realizations of (nominally) the same experiment (model calculation) may show different outcomes. Consider for example the quantity "mean annual wind speed". The value of this parameter in any past year is known (in the sense it can be estimated from observational data) and is a single constant value (except for observational uncertainty). However the mean annual wind speed varies from year to year and hence even if there were no observational uncertainty (i.e. reducible error), the wind speed in any future year would be inherently uncertain.

Fox (1984) presents a convenient framework for considering the various types of modelling uncertainties. Fox indicates that inherent uncertainty (of the type just noted) can be represented by the average of the squared differences between observed values of a quantity and the (time or space) averaged value of the quantity:

$$\langle (C_o - \langle C_o \rangle)^2 \rangle \qquad (2\text{-}1)$$

where

$\langle\ \rangle$ = average value of the quantity
C_o = observed value of a quantity

The magnitude of this term represents inherent uncertainty and hence provides an estimate of the best that any model can hope to do.

Traditionally, little or no attempt has been made to quantify the effects of inherent parameter variability on model predictions. This may stem partly from the practical difficulty of distinguishing between inherent variability, measurement uncertainty and lack of information. The probabilistic analysis concept requires that the random uncertainty associated with each parameter be quantified and a more or less appropriate probability distribution be specified. These uncertainties can then be propagated numerically through the mathematical model resulting in a direct estimate of the uncertainty of the result (see also Chapter 3 and Section 6.2).

Hofer and Krzykacz (1984) also discuss a second class of uncertainty, namely the uncertainty that is inherent to the analysis itself. "They represent possible variability in the analysis of the same realization of the complex and are due to inaccurate knowledge of constants and functional laws." Uncertainties of this type are discussed in the following section (see also Chapter 3 and Section 6.2).

2.3.3 Imprecision in Parameter Estimation

In most applications of an assessment model, parameter information specific to the application is not available. In a more general sense, measurements are always taken under circumstances that are different from the circumstance of interest. Thus, uncertainty is introduced by extrapolating the results of past measurements to a new situation. Fox (1984) notes that "The use of nonrepresentative meteorological data is generally considered to be a major difficulty in regulatory modelling."

In performing the extrapolation, care must be taken to be precise about the definition of parameters. The time scale of the measurements taken and the conditions under which parameter values apply are of particular interest. Considerable uncertainty may be introduced into a model by using data derived from the global literature. This arises because the variability of a parameter over a specific site is likely to be far less than the variability of the same parameter on a much larger scale. Thus, appropriate parameter distributions must be estimated with specific model objectives in mind.

Some attention must also be given to the level of parameter disaggregation required. This aspect of parameter estimation is linked very closely with uncertainties in model formulation. The issue is to realize the level of detail required to produce the desired level of output certainty. For instance, parameter values may be considered separately for distinct soil horizons, distinct species of flora and fauna, hourly, daily or monthly meteorological conditions or alternatively, appropriate lumped parameters may be used in the analysis (see also discussion in Section 2.1).

Uncertainty in parameter estimates can be reduced in several ways. First, expert opinion can be used to derive estimates of parameter distributions for any model application. This subjective process of judging the similarity of parameter estimates and published parameter values and adjusting them appropriately for an application can be formalized by using a likelihood function to quantify similarity and Bayes' theorem to make the adjustment (Henrion and Morgan 1984).

Finally, the matter of appropriate levels of parameter aggregation/disaggregation can be addressed by carrying out the analysis at various levels of parameter aggregation and examining the effect on model results.

2.3.4 Correlations Between Parameters

The problem of selecting appropriate parameter distributions for a particular model application is confounded by the fact that many physical parameters are not independent of each other. Natural interrelationships are quite complex and will never be captured fully by any model or set of parameter distributions. Thus, further uncertainty and error is introduced into model results when two parameters, which are correlated in the real world, are allowed to vary independently in a probabilistic simulation.

There are many types of correlation among parameters. At one extreme is correlation in the uncertainty concerning two truly constant parameters, as might result from statistical analysis of an experiment. At another extreme is the type of uncertainty where two or more quantities are uncertain but if one is assumed known the others become less uncertain. There are many mixed cases.

This source of uncertainty may, in fact, allow the generation of sets of input parameter values which are physically unrealistic or even impossible. For example, if the retardation of a radionuclide is known to increase as the pH of groundwater increases, then a set of input parameter values which included a high groundwater pH and a low radionuclide retardation factor would be inappropriate.

Many of the correlations which exist between parameters in the real world are known with varying degrees of accuracy. Although it is possible to generate correlated parameters (e.g. Iman et al 1980, Kleijnen 1974) or to analytically determine their effect on additive and multiplicative model predictions (e.g. Hoffman and Gardner 1984), it is difficult to obtain enough data to actually estimate covariance terms. Inspection of a variety of models has shown that a few correlations among model parameters can have a significant effect on the variance of model predictions. The greatest changes noted were in the estimates of extreme percentiles in the distributions of the predicted values (Hoffman and Gardner 1983).

Much work remains to be done in this area of sensitivity and uncertainty analysis. Perhaps the most significant contribution will be the development of a formal set of guidelines on how to specify suspected covariances between parameters in the absence of data. One alternative approach is to continue the modelling process, looking for functional relationships among correlated parameters, until the remaining parameters are statistically independent (or nearly so as indicated by the correlation coefficients).

2.4 Some Other Sources of Uncertainty

There are several sources of uncertainty which are difficult to categorize as associated either with the model or with the input parameters. These will be considered briefly here.

Henrion and Morgan (1984) cite disagreement among scientists as a source of uncertainty in mathematical modelling, especially in modelling for radiological assessment. Short term disagreement is common in most areas of science, often revolving around different technical interpretations of the same available information. Usually, the disagreements are resolved and consensus is reached at least in the long run. However, when data are hard to obtain, decisions may have to be made before the technical disagreements are resolved, adding to modelling uncertainty. Technical problems are also encountered when extrapolating beyond the limits of available observations. The fact that individuals may have a "motivational bias" or vested interest in a particular outcome further complicates matters.

Cranwell and Helton (1981) point out that in a probabilistic analysis, there is considerable uncertainty associated with scenarios. A scenario is defined to be a collection of related events, features and processes potentially affecting the situation being modelled. "Scenario uncertainty arises from problems associated with the 'completeness' of scenarios, the definition of parameters which describe scenarios and the rate or probability of scenario occurrence". As long as a single "reference" scenario is considered, scenario uncertainty is not a problem. However, as a probabilistic analysis begins to consider the risk from many scenarios, scenario uncertainty will become a major consideration. The discussions in Sections 3.3 and 3.4 presume that an appropriate scenario has been established. It is also apparent that if the analysis is based on a scenario that is inappropriate in one or more ways then the outcome of the analysis will also be inappropriate.

Another source of uncertainty in modelling effects on human individuals or populations over long periods of time is the uncertainty regarding long term human behaviour (Cameron 1982). This is of considerable concern in modelling the long term environmental effects of high level nuclear waste disposal, since the possibility of human intrusion into an underground repository is finite, unknown and, in our view not amenable to reliable estimation.

2.5 Further Observations Regarding Uncertainty

Despite the fact that the various sources of uncertainty described previously are difficult to examine in isolation, some generalizations on model uncertainties can be drawn, using information from limited model validation studies, stochastic modelling, and the careful use of expert judgement. Hoffman and Miller (1983) conclude that generally, in radiological assessment models:

- uncertainties will usually decrease with increased time and space averaging
- uncertainties are generally less for questions requiring assessments of collective or average risks than for

questions about the risks to individuals or small population groups

The reason for these trends in radiological models is that much of the uncertainty is associated with source term model estimates. Source term uncertainties are unaffected by the spatial averaging that occurs when the question being asked changes the emphasis from one of the radiological impact on individuals to one of the collective impact on larger populations extending over large regions (Hoffman and Miller 1983).

Gardner et al (1982) also make some generalizations regarding the relative importance of various types of uncertainties. They indicate that, "uncertainties in the final predictions are often determined by the model structure and the questions to be addressed by the model. Estimation bias (deviation of predictions from some desired behaviour) may be a minor component in determining the uncertainties of model predictions, yet most models are developed and parameterized to minimize estimation bias." This implies that greater emphasis should be given to minimizing uncertainties associated with the fundamental structure of the model and somewhat less to the calibration or "fine tuning" of the model. This conclusion is similar to that of Fox (1984) who notes: "Although some bias is introduced by the use of meteorologically nonrepresentative observations, one would expect such data errors to average to zero. This is most likely not true for model formulation errors which are likely to introduce bias."

Given the nature and magnitudes of uncertainties associated with mathematical models, there is widespread agreement that models should not be used as the sole determining factor in a pass/fail exercise for decision making. "Models cannot replace rational and informed decision making." (Fox 1984).

3.0 CHARACTERIZING AND PROPAGATING UNCERTAINTY

In general terms the phrase "uncertainty analysis" refers to techniques by which our lack of knowledge of events which might occur and our inability to describe exactly any such events can be translated to probabilistic statements about the effects of any such events. In Chapter 2, factors which contribute to the overall uncertainty in the development and use of models for environmental assessment purposes were discussed. The focus of this chapter is a discussion of techniques for propagating uncertainty associated with the input parameters through mathematical models. First, however, two other important aspects of uncertainty analysis are briefly discussed; namely, scenario analysis and model selection.

3.1 Scenario Analysis

In this report, the term "scenario" is used to mean the collection of information necessary to describe and analyse a particular situation and sequence of events by which radioactivity could be released from a radioactive waste management facility, be transported through the environment and eventually result in human exposure.

Scenario uncertainty arises from questions concerning whether or not all possible scenarios have been identified, the plausibility of particular scenarios and the inability to describe "exactly" these scenarios. The latter source of uncertainty is related to man's limited ability to describe exactly, in a mathematical model, the events and processes which make up a scenario and to define exactly the parameters of the model. These subjects are discussed in later sections of this chapter.

The investigation of systematic methods for identifying possible scenarios, estimating the likelihood of their occurrence and combining the consequences of scenarios is a vast and controversial subject that cannot be covered in this report. Kaplan and Garrick (1981) argue that the often used definition of risk being equal to probability times consequence is misleading. The reduction of a list of multiple scenarios, probabilities and consequences into a single risk parameter leads to loss of information on how to judge the overall acceptability or otherwise of the operation or facility being investigated. For example, a low-probability high-damage scenario (say, a reactor meltdown) and a high-probability low-damage scenario (say, a chronic radioactive emission) with the same probability-consequence factor are clearly not the same thing at all. The resolution of such difficulties often requires non-objective or value-weighted inputs. A study of some of these problems as applied to long-term uranium tailings management can be found in SENES (1985).

3.2 Model Selection

Much of the information presented in section 2.2.1 regarding uncertainties in model formulation is also relevant to any discussion of model selection. Since there will always be differences between mathematical models and reality, a modeller must decide what levels of complexity and uncertainty are acceptable and/or appropriate for a given situation. That is, the important question to answer initially is "How certain do we need to be?" and not "How can we resolve or eliminate uncertainty?"

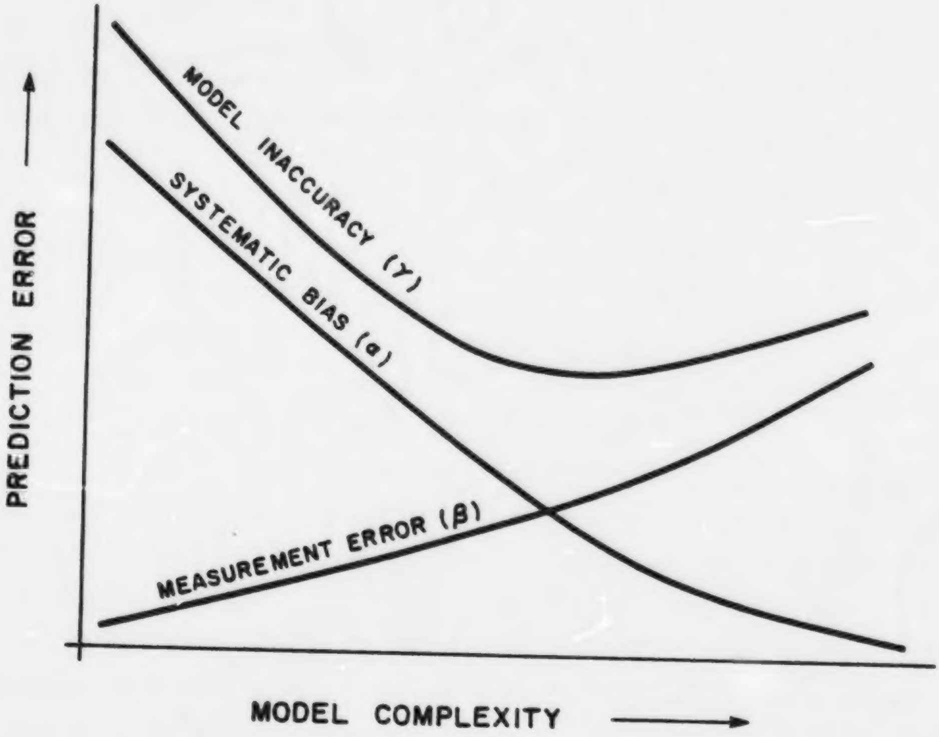
For any situation, a spectrum of mathematical models of increasing complexity will be available. As one moves toward greater complexity, the models increase in resolution, explicitly considering more of the complexities of the real system. O'Neill

(1971) points out that, "This increase in resolution is balanced by a decrease in mathematical tractability and simplicity of interpretation. In addition, the number of model parameters which must be quantified by experiment increases, thereby multiplying the total investment of time and effort." Many authors have pointed out that increased levels of model complexity do not necessarily produce predictions of greater accuracy (Smith and Cohen 1982, O'Neill 1971, Hoffman and Miller 1983).

O'Neill (1971) claims that there is an optimal level of model complexity where the decrease in systematic bias with increasing model complexity is matched by the attendent increase in measurement error. This is illustrated in Figure 3.1. Although the exact shape of the curves must be determined for each specific class of alternate models, the figure illustrates qualitatively how increasing model complexity may, in fact, lead to decreased model accuracy. Given this phenomenon, Hoffman and Miller (1983) find it important to distinguish between assessment models, which are designed to serve as tools for decision making, and research models, which are developed to improve understanding of the function and structure of real systems. Although the increased complexity of a research model offers greater flexibility in terms of the questions it can address, it is likely that the best model for assessment purposes is the simplest model that will produce results with an acceptable degree of accuracy (Hoffman and Miller 1983). Any increase in model complexity is not likely to yield more valid results. As Smith and Cohen (1982) note, "Improvements in the handling of the complexities we do understand in no way compensates for the omission of even a single one of the many complexities we don't know how to handle."

FIGURE 3·1

COMPONENTS OF ERROR IN MODEL PREDICTIONS AS A FUNCTION OF MODEL COMPLEXITY



Note: Systematic bias or alpha error is usually assumed to decrease as models account for more of the complexities of real systems. Measurement or beta error might be expected to increase as more parameters are added to the model. The resultant model inaccuracy, or gamma error ($\gamma = \alpha + \beta$) may, therefore, tend to increase for very complex models. The shape of the three curves and the relationships between them illustrate the hypotheses to be examined in this study. The relationships suggest that there is an optimal level of complexity for a model (from O'Neill 1971).

3.3 Analytic Methods

Given the selection of an appropriate model and estimates of the uncertainties in each of the model parameters, a number of methods are available for propagating the uncertainties in a strict, analytic fashion, through to the model output. That is, the uncertainty in the model output can be expressed as an explicit function of the uncertainties in the input parameters. Experimental science has long used analytic methods to estimate the uncertainty derived in a quantity (the model output) that is a function of one or more experimental and uncertain parameters. These analytic methods of uncertainty analysis include the techniques used for error propagation and include the well-known propagation-of-error formula. If a quantity, x, is a function of independent variables u, v, w,... given by

$$x = f(u, v, w, \ldots) \tag{3-1}$$

then the estimated variance on x is given approximately by

$$s_x^2 = \left(\frac{\delta f}{\delta u}\right)^2 s_u^2 + \left(\frac{\delta f}{\delta v}\right)^2 s_v^2 + \left(\frac{\delta f}{\delta w}\right)^2 s_w^2 + \ldots$$

$$+ \text{ covariance terms} \tag{3-2}$$

where ${s_u}^2$, ${s_v}^2$, ${s_w}^2$, ...are the estimated variances on u, v, w..., respectively. A simplified derivation of equation (3-2) is given in Taylor (1982). The partial derivatives are evaluated with all variables held at their mean values. This formula is usually used with the assumption of normal distribution functions in order to estimate confidence intervals about the mean of the output value x (i.,e. the shape of the distribution of x has to be known) but in fact the formula applies to any form of distribution of u, v, w, ... The formula is based on the first

TABLE 3.1

Propagation-of-Error Formulas for Some Simple Functions

Function Form*	Approximate variance	Additional term if u and v are correlated
$x = au \pm bv$	$s_x^2 = a^2 s^2_u + b^2 s_v^2$	$\pm 2ab s_{uv}$
$x = auv$	$\frac{s_x^2}{\bar{x}^2} = \frac{s_u^2}{\bar{u}^2} + \frac{s_v^2}{\bar{v}^2}$	$+2 \frac{s_{uv}}{\bar{u}\bar{v}}$
$x = \frac{au}{v}$	$\frac{s_x^2}{\bar{x}^2} = \frac{s_u^2}{\bar{u}^2} + \frac{s_v^2}{\bar{v}^2}$	$-2 \frac{s_{uv}}{\bar{u}\bar{v}}$
$x = u^2$	$s_x^2 = 4 \bar{u}^2 s_u^2$	
$x = \sqrt{u}$	$s_x^2 = \frac{s_u^2}{4\bar{u}^2}$	
$x = au^{\pm b}$	$\frac{s_x^2}{\bar{x}^2} = b^2 \frac{s_u^2}{\bar{u}^2}$	
$x = a \exp(\pm bu)$	$\frac{s_x^2}{x^2} = b^2 s_u^2$	
$x = cu^{\pm a} v^{\pm b}$	$\frac{s_x^2}{\bar{x}^2} = a^2 \frac{s_u^2}{\bar{u}^2} + b^2 \frac{s_v^2}{\bar{v}^2}$	$2(\pm a)(\pm b) \frac{s_{uv}}{\bar{u}\bar{v}}$
$x = \ln(u)$	$s_x^2 = \frac{s_u^2}{\bar{u}^2}$	
$x = a \ln(\pm bu)$	$s_x^2 = a^2 \frac{s_u^2}{\bar{u}^2}$	

*a,b and c are constants; u and v are assumed to be statistically independent; and the value of x is finite and real [i.e. v is non-zero for ratios with v as denominator, u>0 for u and ln(u), and ($\pm$bu)>0 for ln($\pm$bu)].

order Taylor series expansion of x about the mean of x and assumes that the partial derivative terms of higher order are negligible. (See Ku (1966) for more detail.)

For the simple case of x= u + v, the propagation-of-error formula leads to

$$s_x^2 = s_u^2 + s_v^2 + 2s_{uv}$$

where s_{uv} is the covariance of u and v. For uncorrelated input variables ($s_{uv} = 0$), the variance on x is just the sum of the variances on u and v. Propagation-of-error formulas derived from equation (3-2) for other simple functions are given in Table 3.1 (Collee et al 1980). In principal, the propagation-of-error formula can be used for any model function and any number of parameters.

Equation (3-2) shows that, for uncorrelated variables, decreasing the variance of one or more of the input variables will always decrease the variance of the output function x. This can be seen for example in the case of determining the value of a parameter x that is a function of several, independent experimentally measured parameters, u, v, w... Obviously, the more precisely u, v, w... are known (the smaller their variances), the more precisely x is known (the smaller its variance).

This does not apply to correlated parameters. Consider the sum of two positively correlated random variables u and v. The more that u and v are correlated, the more they would tend to take on high values together, or low values together, leading to higher or lower sums, respectively. Thus, the variance of the output, the sum of u and v, would increase. On the other hand, a negative correlation between u and v would lead to compensating effects and a lower variance in their sum. For functions more complex than a summation of variables, no general statement can be made concerning how changing the variances of correlated input variables will affect the variance of the output function.

The approach of equation (3-2) to propagating uncertainty through the model is limited to the estimate of the variance of the model output - it says nothing about the shape of the output function. The relative "degree of uncertainty" in x can be estimated as $s_x/\overline{x}^2$ if further inferences on the uncertainty in x are to be made (e.g. confidence intervals about the mean of x or the probability that $x \leq$ some limiting value), the shape of the output function x must be known.

To overcome this difficulty, it is often assumed that the input variables u, v, w... are normally distributed, and for simple additive ($x = u + v + w + \ldots$) models (with normal input variables) or multiplicative ($x = u.v.w....$) models (with lognormal input variables), the output function x is also distributed normally or lognormally, respectively. As a consequence of the statistical processes as stated in the Central Limit Theorem of Statistics, this is often the case even when the input parameters are not normally distributed. The output function x will tend to approximate a normal (or lognormal) distribution if the number of parameters being added (or multiplied) is sufficiently large and the variance of any parameter is negligible compared to the sum of the variances of the parameters.

Figures 3.2 to 3.6 illustrate the Central Limit Theorem for a simple additive model with from one to five random input variables drawn from uniform distributions. This and other models are discussed in more detail in Section 3.4.9 and in Chapter 4.

FIGURE 3·2

HISTOGRAM OF YI = A + B + C + D + E WITH VARIABLE A UNIFORMLY DISTRIBUTED FROM 1 TO 3 AND B = C = D = E = 2·0

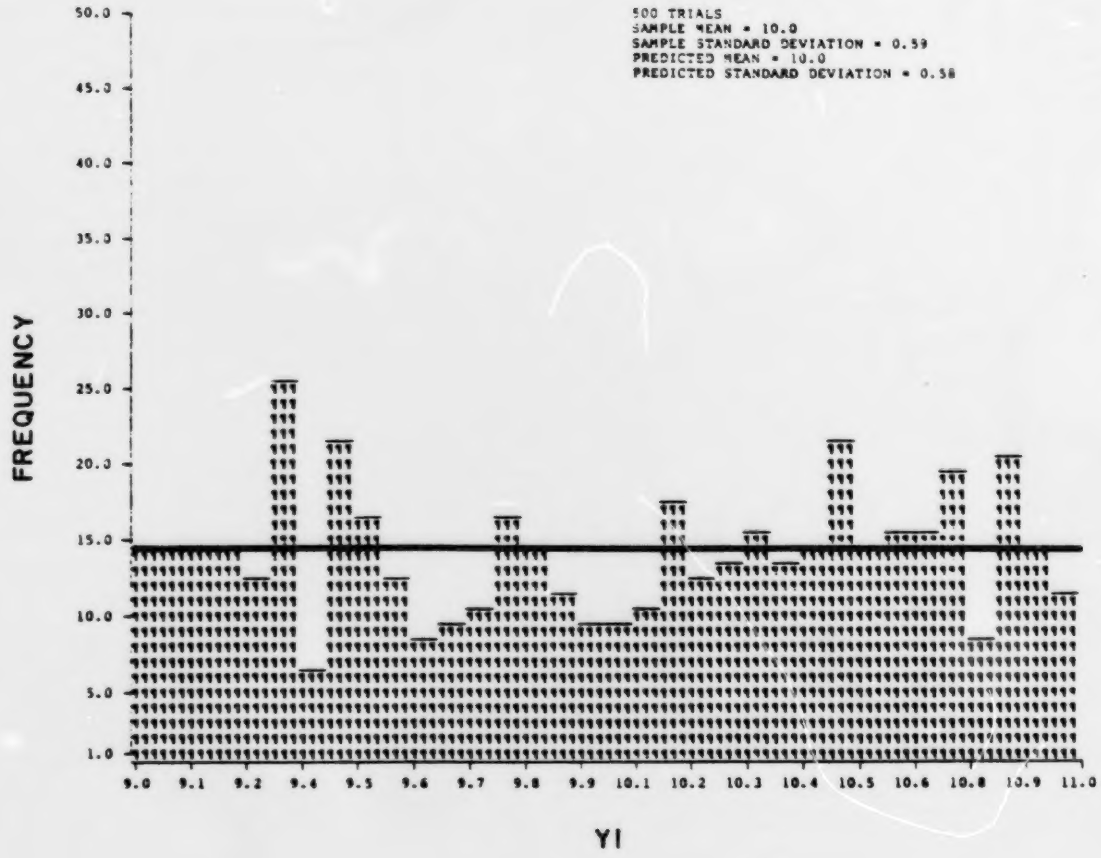


FIGURE 3·3

HISTOGRAM OF YI = A+B+C+D+E
WITH A AND B UNIFORMLY DISTRIBUTED FROM 1 TO 3
AND C=D=E=2·0

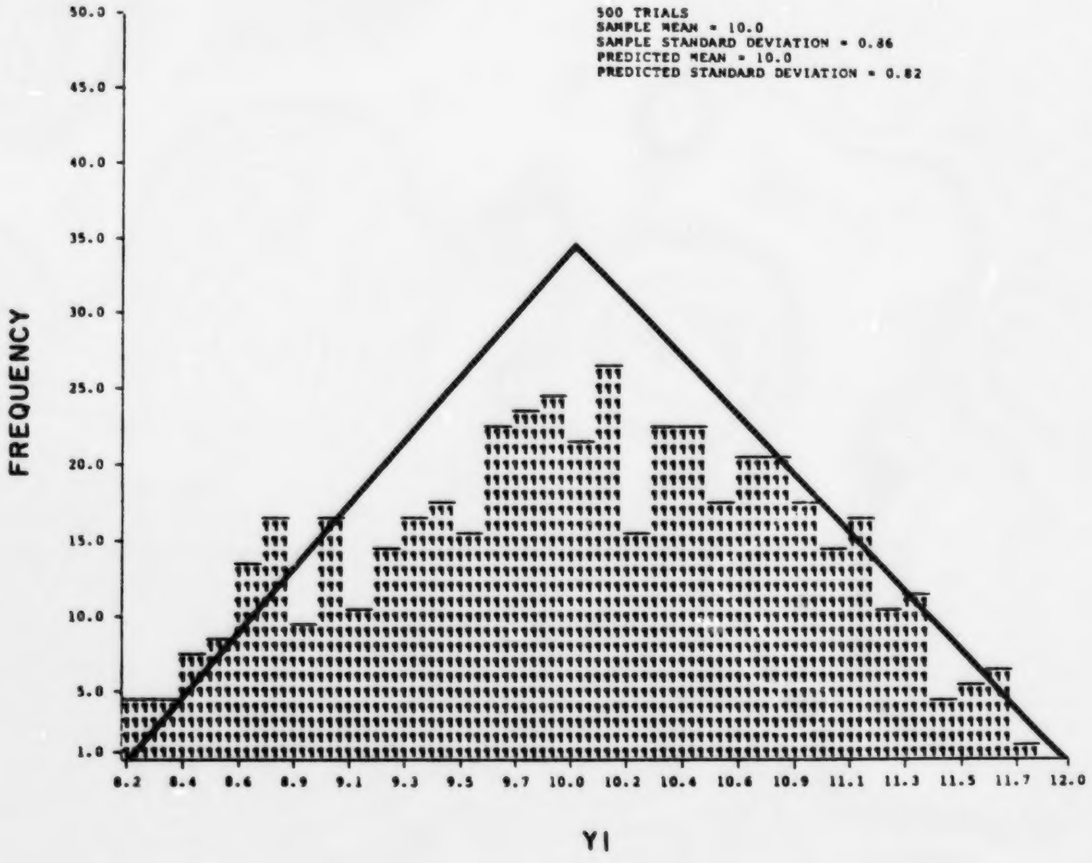
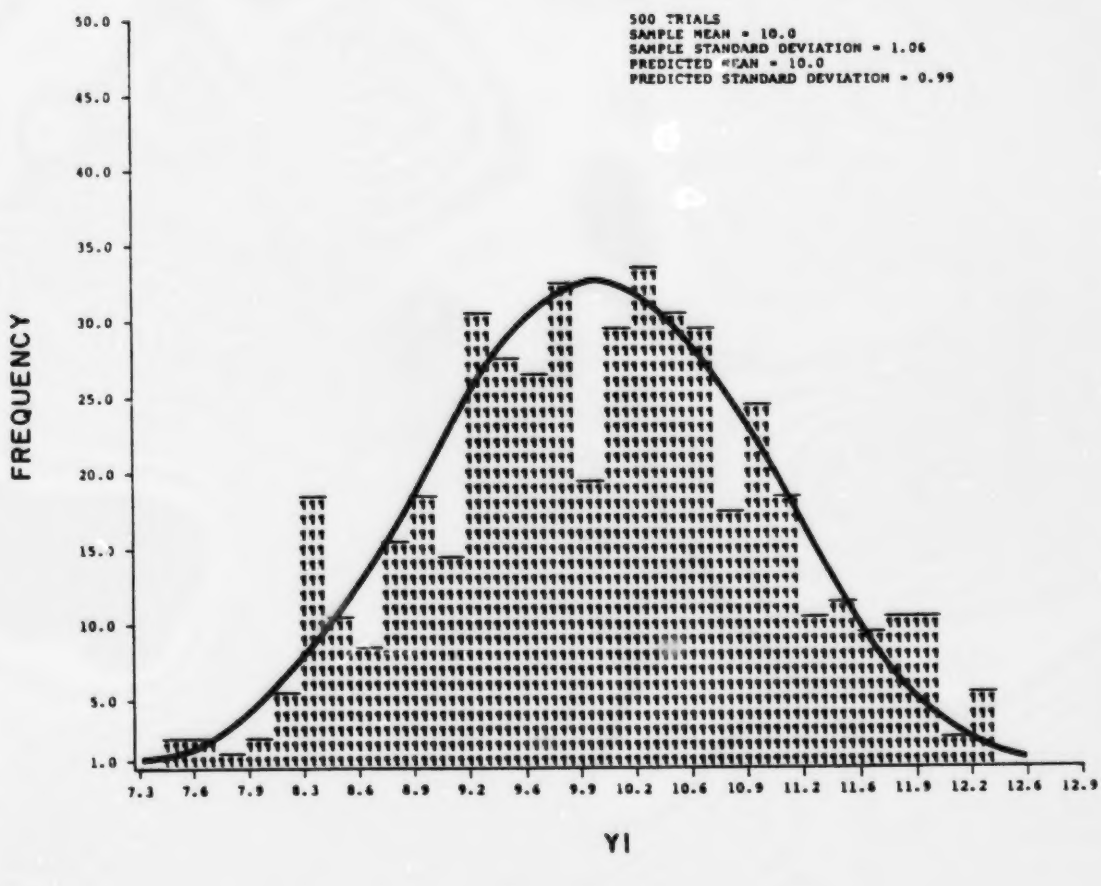


FIGURE 3·4

HISTOGRAM OF YI = A+B+C+D+E
WITH A, B, AND C UNIFORMLY DISTRIBUTED FROM 1 TO 3
AND D = E = 2·0



500 TRIALS
SAMPLE MEAN = 10.0
SAMPLE STANDARD DEVIATION = 1.06
PREDICTED MEAN = 10.0
PREDICTED STANDARD DEVIATION = 0.99

FIGURE 3·5

HISTOGRAM OF YI = A + B + C + D + E
WITH A,B,C AND D UNIFORMLY DISTRIBUTED FROM 1 TO 3
AND E = 2·0

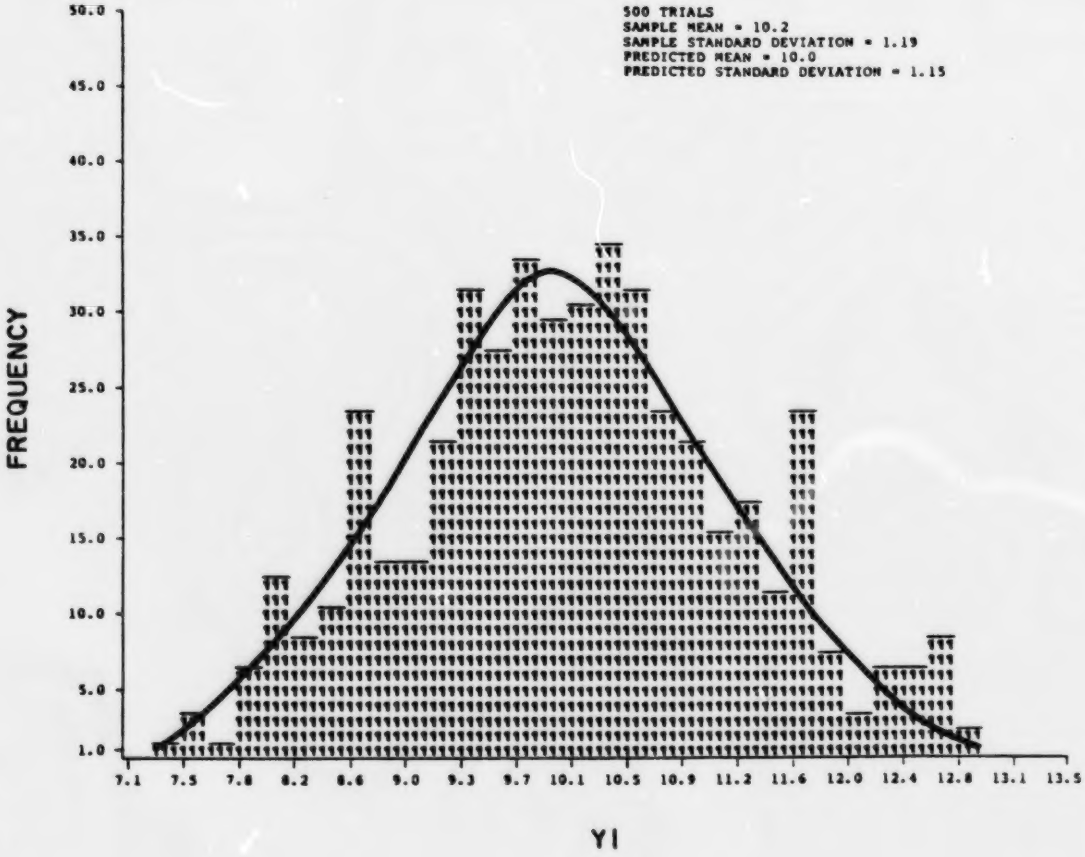
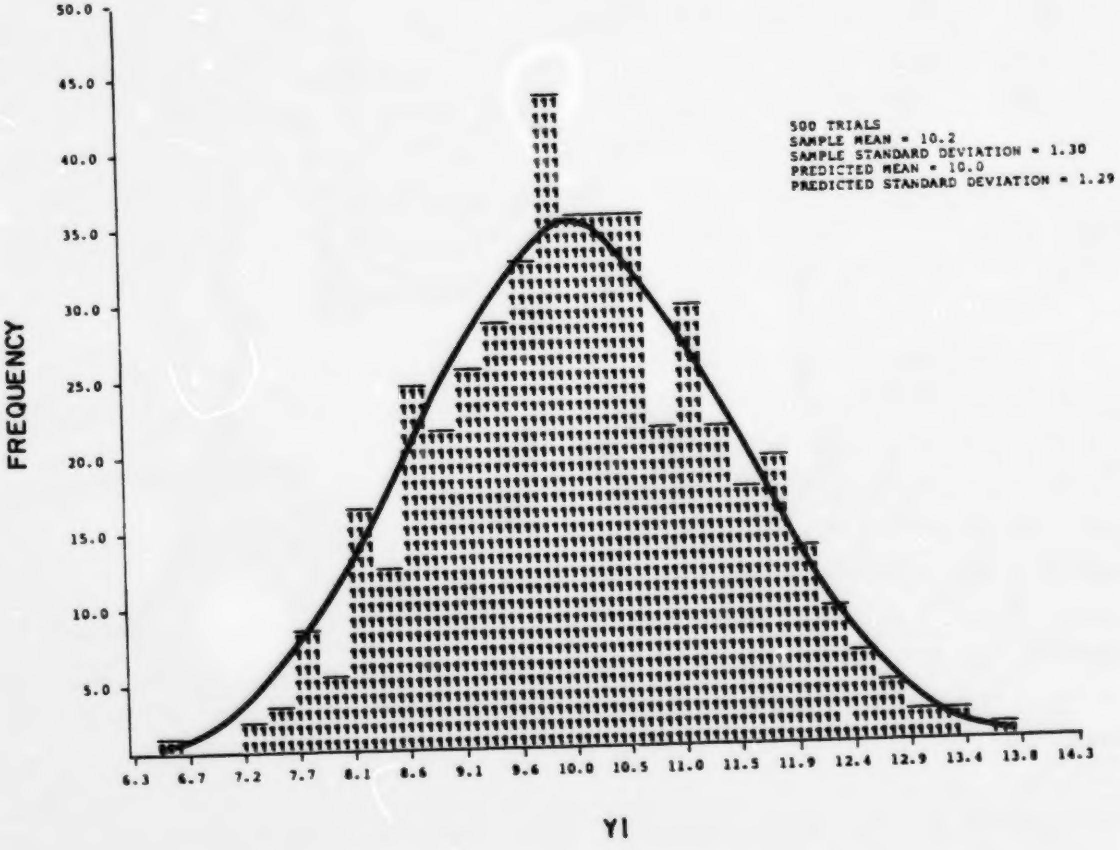


FIGURE 3·6

HISTOGRAM OF YI = A+B+C+D+E
WITH ALL INPUTS UNIFORMLY DISTRIBUTED FROM 1 TO 3



The figures show in each case that the standard deviation (square root of the variance) of the output distribution predicted by equation (3-2) is in good agreement with the calculated sample standard deviation. Furthermore, as the number of input variables increases from one to five (Figure 3.2 to Figure 3.6), the shape of the output distribution tends to approximate more accurately a normal distribution even though the input distributions are each uniform, (i.e. "flat") and far from normal. The curves superimposed on the output histograms in Figures 3.4, 3.5 and 3.6 are normal distributions based on the predicted means and variances of the output. The curve in Figure 3.3 is the predicted triangular output from summing two uniformly distributed variables. The curve superimposed on Figure 3.2 is the predicted uniform distribution from the sum of four constants and one uniform distribution.

The complete analytic solution to the propagation of uncertainty in any model basically requires answering the question: If the model output x is a function of random variables u, v, w,... with known distribution or probability densities p_u, p_v, p_w,..., what is the probability density p_x? There are many models which are more complicated functions than simple addition or multiplication and which have many parameters. Exact analytic methods for defining the distribution of x for increasingly complicated models rapidly become extremely difficult, if not impossible, to use.

The problem is essentially that of the transformation of variables in probability space. When considering the relatively simple case of functions of a single random variable [x = f(u)], it can be shown (Beckmann 1968) that

$$p_x = p_u \, [f^{-1}(x)] \, |du/dx| \qquad (3\text{-}3)$$

where p_x and p_u are the probability density functions of x and u, respectively (See also Appendix A).

The application of equation (3-3) can be seen with the use of a few simple examples. If a voltage v that is uniformly distributed between 0 and 2 volts is applied across a resistor r, what is the distribution of the power, $w = v^2/r$? For this case, we have

$$p_v = 1/2 \quad (0<v<2) \tag{3-4}$$

Also, since

$$v = \sqrt{wr} \quad \text{and} \quad \left|\frac{dv}{dw}\right| = \frac{\sqrt{r}}{2\sqrt{w}} \tag{3-5}$$

then from equation (3-3)

$$\begin{aligned} p_w &= p_v(\sqrt{wr}\,) \cdot \frac{\sqrt{r}}{2\sqrt{w}} \\ &= \frac{1}{2} \cdot \frac{\sqrt{r}}{2\sqrt{w}} \\ &= \frac{1}{4}\sqrt{r/w} \quad (0 \leq w \leq 4/r) \end{aligned} \tag{3-6}$$

Note that although the probability density p_w becomes infinite at $w = 0$, the probability $p_w dw$ remains finite; in fact, by the definition of a probability density function, p_w must be non-negative for all w and $p_w dw$ must integrate to unity.

Consider a second example. If $x = \exp(u)$ where u is normal with mean a and standard deviation σ, what is the probability density p_x? We have by definition

$$p_u = \frac{1}{\sqrt{2\pi}\sigma}\exp[-(u-a)^2/2\sigma^2] \tag{3-7}$$

Therefore, in equation (3-3), $f^{-1}(x) = u = \ln(x)$ and

$$p_u[f^{-1}(x)] = \frac{1}{\sqrt{2\pi}\,\sigma} \exp[-(\ln(x)-a)^2/2\,\sigma^2] \qquad (3\text{-}8)$$

Also $|du/dx| = 1/x$ and therefore from equation (3-3)

$$p_x = \frac{1}{x\sqrt{2\pi}\,\sigma} \exp\,[-(\ln(x)-a)^2/2\,\sigma^2] \qquad (3\text{-}9)$$

Thus, the variable x is lognormally distributed.

The examples above treat functions of only one random variable. For multidimensional distibutions, which are of interest in environmental models, the situation often becomes complex. For example, the product $x = uv$ of two independent random variables u and v has a probability density given by (Beckmann 1968)

$$p_x = \int_0^{\infty} p_u(u)\; p_v\left(\frac{x}{u}\right) \frac{1}{u}\, du - \int_0^{-\infty} p_u(u)\; p_v\left(\frac{x}{u}\right) \frac{1}{u}\, du \qquad (3\text{-}10)$$

For the simple case of u and v each being uniformly distributed over the range (0,a), equation (3-10) leads to

$$p_x = \frac{1}{a^2} \ln\left(\frac{a^2}{x}\right) \qquad (0 \leq x \leq a^2) \qquad (3\text{-}11)$$

a rather unexpected result!

Other than for very simple functions, the exact analytic solution to the determination of the probability density of the model output x rapidly becomes very difficult or impossible to find. Integral transform methods (e.g. Fourier, Laplace, Mellin) can be used but again are only practical for simple functions (Seiler

1984). Approximation methods such as the Method of Moments might be workable in some cases. In this method, rather than determining the exact shape of the output distribution from the input distribution, the moments of the output distribution are calculated from the moments of the input distributions. This was discussed earlier for the first (mean) and second (variance) moments (equation 3-2) but it can be extended to higher moments with moment generating functions (Beckmann 1968). Higher moments give more detail about the shape of the output distribution and in the limit, after all the moments are known, the entire shape of the output is known. However, again, only simple models can be accommodated by this method.

Not even considered in the examples above is the added complexity of correlated input variables. Thus, statistical methods such as those which involve Monte Carlo sampling techniques are usually the only practical tools to use in propagating uncertainty through other than very simple models.

3.4 Simple Random Sampling and Alternatives

3.4.1 Introduction

The objective of this section is to discuss random sampling, and its alternatives, and the advantages and disadvantages associated with the various methods. Furthermore, a comparison of the major sampling methods - unstratified or simple random sampling, and various types of stratified random sampling including latin hypercube sampling - for use in Monte Carlo modelling is made. The review presented in this section represents opinions found in the literature. In preparing this survey, the eventual application, that is, the selection of values of input variables for Monte Carlo studies, was kept in mind, and hence while a broad range of sampling methods was reviewed, only certain pertinent ones are discussed here. As indicated below, no generally conclusive answer was found as to which is the best method to be used since the most appropriate method is determined by the desired end use of the analysis. However, as noted in section 3.4.8, simple random sampling and latin hypercube sampling are the methods suggested for application in probabilistic modelling.

3.4.2 Simple Random Sampling (RS)

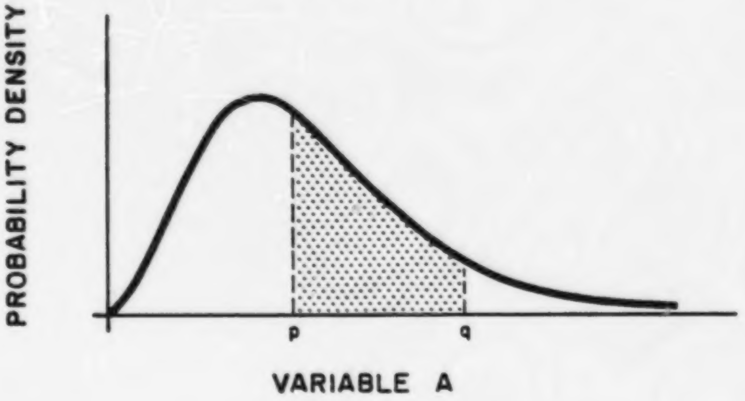
Raj (1968) gives the following definition of simple random sampling: "From a population of N units select one by giving equal probabilities to all units. Make a note of the unit selected and return it to the population. Perform this operation n times to get a simple random sample of n units selected with replacement." Note that from the point of view of the random variable, this implies that each possible value is chosen according to its probability. That is, if x_1 is twice as probable a value for the random variable as x_2, then the population from which the sampling is drawn will have twice as many members with value x_1 as with value x_2.

In Figures 3.7 and 3.8, RS is illustrated for a situation in which there are two random variables A and B. If a total of 16 sets is required, 16 independent observations are made on each of A and B. The sample consists of 16 pairs, each pair being made up of an observation on A and an observation on B.

The graphs in Figure 3.7 are probability density functions. They have the property that areas under the curve represent probabilities; that is, in each case the hatched area represents

FIGURE 3·7

SIMPLE RANDOM SAMPLING (R.S.)



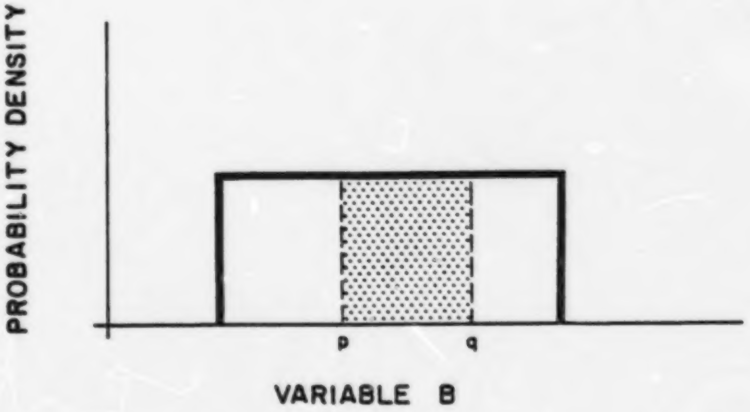
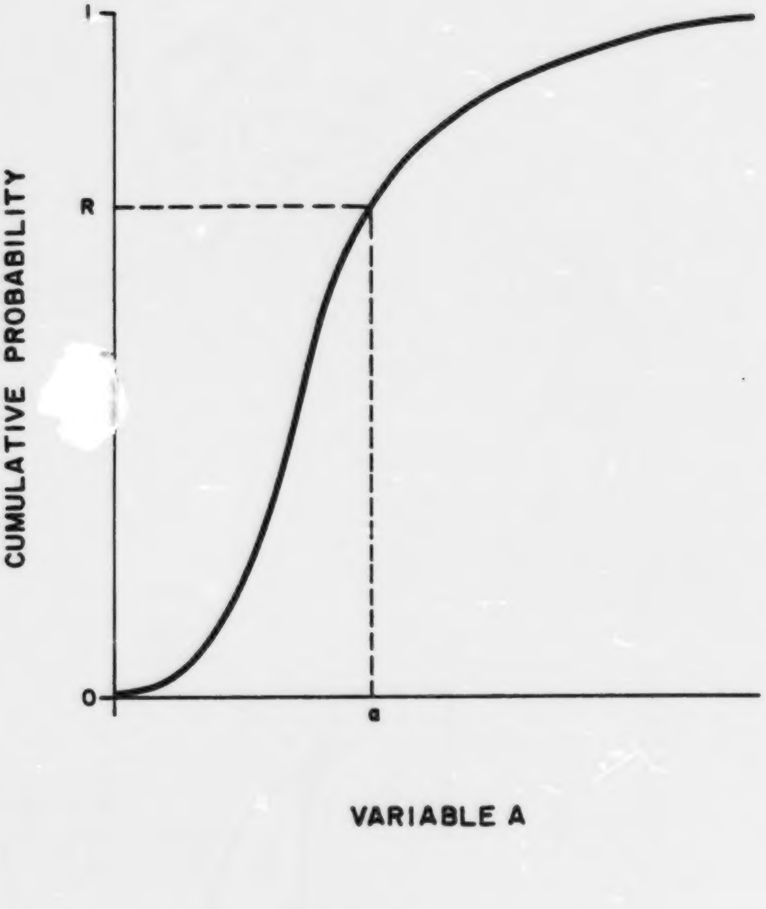


FIGURE 3·8

SIMPLE RANDOM SAMPLING
CUMULATIVE PROBABILITY FUNCTION FOR A



the probability that the random variable (A or B) will take on a value between p and q on being randomly sampled. Thus if f(x) is the probability density function of a random variable x, then,

$$Pr(p \leq x \leq q) = \int_{p}^{q} f(x)\,dx \qquad (3\text{-}12)$$

The graph in Figure 3.8 is the cumulative distribution function for the random variable A. Its value for a given a is numerically equal to the probability that the random variable A will be equal to or less than a. If F(a) is the cumulative distribution function and f(a) the probability density function for A, then

$$F(a) = \int_{-\infty}^{a} f(x)\,dx \qquad (3\text{-}13)$$

The observation on B is made simply by choosing a uniformly distributed random number from the appropriate range (i.e. from between the endpoints of B) since as seen in Figure 3.7, B is uniformly distributed. To make a single random observation on A, the cumulative distribution function as shown in Figure 3.8 is used. A value R of a uniformly distributed random variable on the range 0 to 1 is chosen, for example by a computer. The corresponding point on the vertical scale of Figure 3.8 is found, projected over to the graph as shown, and down to the horizontal scale to provide a value for A, namely a. It can be shown that random numbers so chosen are distributed according to the probability density function for A shown in Figure 3.7. This and other methods of sampling from a random distribution are discussed in Appendix A.

An advantage of simple random sampling (i.e. unstratified sampling) is that it is the most simple of all of the ordinary methods of sampling. It is elementary to obtain unbiased estimates of the mean and variance, and the frequency distribution of a sample converges to the probability density function as the sample size increases. Detailed computer programs or large computer memory is not needed. A disadvantage is that a large number of samples (and hence computer time) may be needed to estimate statistics as the desired confidence level (see section 3.4.7).

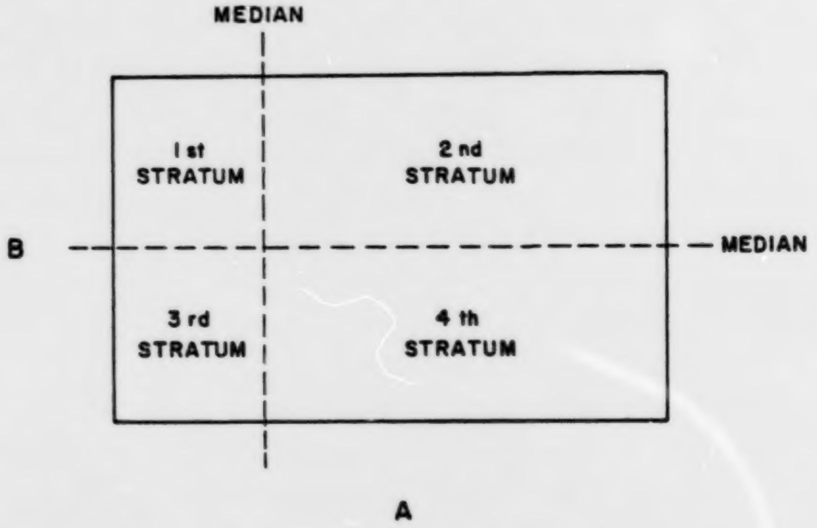
3.4.3 Stratified Random Sampling (SRS)

McKay et al (1979) explain that in SRS, all areas of the sample space S of X are represented by input values. The sample space S is partitioned into I disjoint strata, S_i, where $p_i = P(x \varepsilon S_i)$ represents the size of S_i. If I=1, this corresponds to RS. A random sample x_{ij}, $j=1,\ldots,n_i$ from S_i is then obtained such that $\Sigma n_i = N$. This explanation of SRS is somewhat deficient in that the values of I (the number of strata) and of the n_i's are not specified. Presumably, $n_i = p_iN$ and $\Sigma n_i = N$.

Consider the example described for RS. The ranges of values of both A and B are divided into subranges by any desired method. Suppose the ranges of A and B are each divided into two parts at the median, as illustrated in Figure 3.9. Then if a total sample of 16 sets of values is required, four pairs of observations (a pair is one observation on each of A and B) are made by simple random sampling (RS) in each of the quadrants (i.e. strata) of Figure 3.9. Since the ranges of the variables are split at the median and since proportional (to the probability) stratification would likely (but not always) be used, each stratum has equal sampling in this case. Note that the ranges of A and B are not necessarily split at any particular point. Here they are split at the median.

FIGURE 3·9

STRATIFIED RANDOM SAMPLING (S.R.S.)


MEDIAN
1st STRATUM
2nd STRATUM
B
MEDIAN
3rd STRATUM
4th STRATUM
A

Stratification may produce a gain in precision in the estimate of characteristics, such as the mean, of the whole population. It may be possible to divide a heterogeneous population into subpopulations which are internally homogeneous with respect to a particular characteristic of interest. If each stratum is homogeneous, precise estimates of any stratum mean can be obtained from a small sample. These precise estimates can then be combined to get a precise estimate for the whole population. In general then, one would expect smaller variances for estimators from SRS than from RS but this is not always true, particularly if the sample size for each stratum is far from the optimum (Cochrane 1977). The optimal stratum sample size is proportional to the product of the total number of units in the stratum and the variance within the stratum.

3.4.4 Quota Sampling (QS)

In this approach to sampling, which has been widely used in opinion and market research surveys, the sample size required in each stratum is computed in advance so that stratification is proportional to the specified quotas. Common variables for stratification in this type of problem are geographic area, age, sex and some measure of income level. The person carrying out the sampling must then sample, in a particular geographic area say, until the set quotas are filled. If strict SRS were used, a great deal of sampling would be necessary to fill the quotas since in the latter stages, many samples chosen would fall into quotas already filled. The filling of quotas is expedited by allowing for some nonrandom selection procedure. In other words, quota sampling can be viewed as stratified sampling with more or less nonrandom selection of units within strata. This can of course produce samples which are biased with respect to certain population characteristics.

3.4.5 Systematic Sampling (SYS)

This sampling method is quite different from simple random sampling. Cochrane (1977) gives the following explanation of systematic sampling. First the N units in the population are numbered 1 to N. Then a unit is chosen at random among the first k. Every kth unit is chosen thereafter to get a sample of size n.

The apparent advantages over simple random sampling are twofold:

1. It is easier to draw a sample.
2. It is intuitively more precise, about as precise as SRS with one sample unit per stratum with the units occurring at the same relative position. The systematic sample is more evenly spread out over the population and can therefore be considerably more precise than SRS.

A variant of systematic sampling involves selecting each unit at or near the center of each set of k units, which may be considered a stratum. There are grounds which indicate that this is more precise than SYS initiated by choosing randomly among the first k units.

Systematic sampling is related to cluster sampling, which is not commonly used in sampling for Monte Carlo purposes. In it, the population is divided into sets called clusters and one or more entire clusters is chosen at random. In systematic sampling, the cluster consists of the first element of the sample, along with every kth member of the population after it. A systematic sample therefore is a simple random sample of one cluster unit from a population of k cluster units.

SYS is more precise than RS if the variance within the systematic samples is larger than the population variance as a whole. SYS is considered to be less precise than SRS. The performance of SYS is dependent upon properties of the population.

Another variation of SYS would be stratified systematic sampling. This is achieved by obtaining a systematic sample within each systematically chosen stratum. This method will be more precise than SRS if systematic sampling within strata is more precise than simple random sampling within strata.

The disadvantages of SYS are that this method gives poor precision when unsuspected periodicity is present and that no trustworthy method is known for estimating the variance of the mean estimates from the sample data. The method of Discrete Probability Distributions (DPD's), which has some of the flavour of SYS, is described in detail in Appendix B.

3.4.6 Latin Hypercube Sampling (LHS)

LHS is also based on SRS and is related to quota sampling, lattice sampling and to the random balance experimental designs discussed by Satterthwaite (1959). LHS takes the idea of SRS a step further in that not only is the full sample space of all variables represented, but each input variable has all portions of its distribution sampled. Suppose there are p random variables X_i, i = 1,2..., p, to be sampled to obtain a sample of size N, each element of the sample consisting of one observation on each of X_i, i = 1,2,...,p. Then, the range of each variable X_i is divided into N intervals of equal probability, an observation is made in each interval, and the N values so obtained for each variable are matched by random choice with the values of the other variables. This procedure can be viewed as a p-dimensional extension of that required to produce a Latin

Square experimental design. The Latin Square experimental design is a technique (essentially randomization subject to a double restriction) which ensures that all of the experimental treatments (under two experimental conditions, hence the term square) to be compared are equally represented.

Consider once again the example used in Sections 3.4.2 and 3.4.3. To obtain a sample of size 16 on A and B, the ranges of A and B are broken up into 16 intervals of equal probability as shown in Figure 3.10. The intervals are then paired by a random process, for example by writing down the numbers from 1 to 16 and next to them a random permutation of the same numbers.

A:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
B:	10	9	4	7	8	11	14	15	3	2	16	13	5	1	6	12

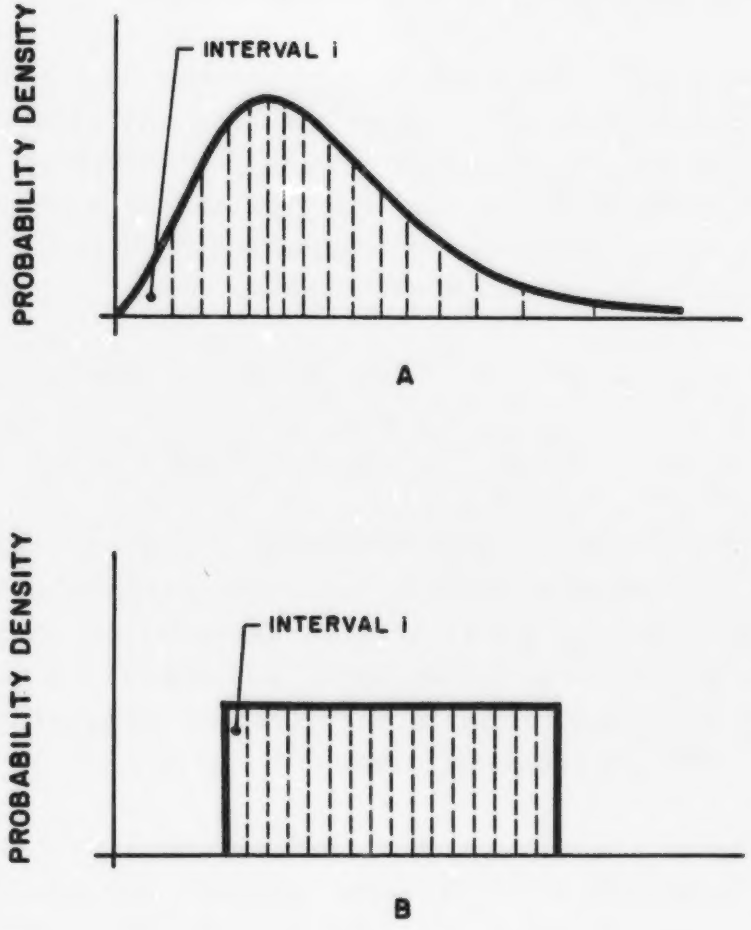
Then the sample is obtained by making a random observation in each interval of each pair by RS. As previously stated, the sample consists of 16 pairs of observations, one from A and one from B in each pair. The first pair would consist of an observation taken at random in the first interval of A along with one on the 10th interval of B, etc.

SRS and LHS are similar in that the low and high values of each variable are combined in the same manner with the low and high value of every other variable. In SRS, this is done probabilistically; in LHS, it has a systematic component in that each variable is sampled in every one of the intervals in its range.

It is interesting to note that while McKay et al (1979) divide the range of each variable into N intervals or strata of equal probability, Helton and Iman (1982) divide the input into N

FIGURE 3·10

LATIN HYPERCUBE SAMPLING (L.H.S.)


PROBABILITY DENSITY
INTERVAL i
A
PROBABILITY DENSITY
INTERVAL i
B

NOTE:

VERTICAL LINES SHOWN ON PROBABILITY DENSITY FUNCTION ARE TO INDICATE EQUAL AREAS UNDER THE CURVE (NOT TO SCALE) intervals of either equal probability or equal width. The latter is not advisable since it implies that all intervals are equally likely which they are not (except for uniformly distributed values). They should be weighted according to their probability. Subdivision on the basis of equal width will lead to a biased estimate of the mean and variance. In their "Latin Hypercube Sampling Program Users Guide", Iman et al (1980) suggest division according to probability only.

3.4.7 Comparison of Sampling Methods with Respect to Monte Carlo Calculations

In this section, a comparison between random sampling (RS), stratified random sampling (SRS) and latin hypercube sampling (LHS) for selecting values of input variables, is made. The discussion is primarily based on the papers by McKay et al (1979) and Iman and Conover (1980).

McKay et al (1979) report the following results for a direct comparison of the three methods. First, all three methods give comparable, unbiased estimates of the sample means, $\overline{Y}$. The estimated standard deviations of the means show that $SD(\overline{Y}_{LHS}) < SD(\overline{Y}_{SRS}) < SD(\overline{Y}_{RS})$ suggesting that LHS yields the most precise estimate (in the sense of minimum variance). As pointed out by Easterling (Iman et al 1980), however, one serious drawback with most studies of LHS is that one cannot tell if LHS outperformed RS since the variance of the estimator cannot be estimated from a single LHS run. The remedy is to run LHS for r runs of size N/r rather than a single LHS run of size N. The latter method was used by McKay et al (1979) to obtain the above comparison. In addition, McKay et al determined that all three methods gave comparable estimates of the mean of the population variance. However, once again LHS resulted in a more precise estimate. Finally, McKay et al report that all three methods yield comparable estimates of the mean of the cumulative distribution function but, as before, LHS yields a better estimate. They conclude that LHS improves upon SRS which is better than RS.

Lawrence (Iman et al 1980) points out that LHS must be used with caution since it may in fact give poorer estimates, in the sense of having larger variances, if the output is not monotonic. Monotonicity here is meant with respect to the underlying function which relates X to Y (i.e. Y=Y(X)). It must be continuous and monotonically increasing or decreasing. This is of course hard to determine in advance since the function can be quite complicated.

Iman and Conover (1980) state that not all models allow for straightforward application of LHS, again citing the monotonicity conditions. In LHS, the variance of the estimators is usually smaller than the variances of estimators obtained from other schemes but this result may be closely related to the monotonicity property of the code. Population variance estimates do have some associated bias. However Iman and Conover claim that if the variance of estimators such as the mean is reduced this bias will be small.

Iman and Conover (1980) also claim that a very useful feature of LHS is that it allows one to measure sensitivity of the code to some of the assumptions of the model behind the code. In particular, after a set of runs is completed under certain distribution assumptions on the input variables, the assumed distribution functions may be altered and an estimate of the corresponding change in the distribution of the output variable may be made without making additional runs of the code. This would be very useful since these codes often require a great deal of computer time to run. If the distribution function of the output variables is altered significantly by the adjustments to the input distribution then the output may be considered to be sensitive to assumptions regarding the input distribution. McKay and Beckman (Iman et al 1980) disagree with changing the input distribution after the run has been made without rerunning the code. They claim that to make inferences regarding the sensitivity of output variables to these changes, the price of rerunning the code must be paid. The concern of McKay and Beckmann (Iman et al 1980) arises since it is not clear if errors made in misspecification of the input distribution will not add up to be significant when obtaining mean estimates, even though the effect of misspecification may be small for each specific cell. The misspecification is the result of using exactly the same values of the variables over again, when in principle they should be re-randomized, at least within the chosen intervals.

Finally, Easterling (Iman and Conover 1980) points out that as a design for sensitivity analysis, LHS is an example of the pure random balance design proposed by Satterthwaite (1959). It is an $N^{-(p-1)}$ fraction of an N^p design, where p is the number of variables, the particular choice and combination of variables being determined by random selection. These designs when proposed were heavily criticised (see discussions following Satterthwaite (1959) by Youden, Kempthorne, Tukey, Box and Hunter). Easterling fails to see why such an inefficient design should be used for that particular type of problem. Recent unpublished work by both the present authors and others (Andres 1985) supports the generally held view that LHS gives values of means of functions of the input variables which have slightly to markedly smaller variance than RS, depending on the form of the function. Simple sums of the input variables tend to show much better performance by LHS while functions with sharp "spikes" show LHS to be only marginally better.

3.4.8 Suggested Sampling Methods

Given the information just presented, only two sampling methods, simple random sampling (RS) and latin hypercube sampling (LHS), should be considered for applications in probabilistic analysis. The former is easier to understand and to implement while the latter appears to give somewhat more consistent results in most cases. The final choice of sampling method depends on the desired use. If the mean of output distributions under probabilistic assessment is required, relatively few trials are implied and the larger computer storage requirements of LHS can be supported by the usually improved variance of the estimated means. However if the full output distribution is required, many more trials are needed and thus RS may to preferred. Obviously there are many cases where there is no clear-cut preference between methods.

3.4.9 Some Examples

In order to illustrate, in a quantitative manner, the comparison between latin hypercube sampling (LHS) and simple random sampling (RS), a number of example "models" were developed (Table 3.2). The models, which were also used in examining sensitivity and uncertainity analysis techniques (Chapter 4), were designed to be mathematically simple so that the characteristics of the methods under examination would not be lost in model detail and so that the expected performance of each model could be calculated exactly by analytical methods. Furthermore, the use of consistent units in the example models is not a problem. Nonetheless, the models do contain much of the flavour of environmental models in that they range from simple additive (e.g. summation of dose from various pathways) and multiplicative models (e.g. ingestion dose = intake rate x contaminant concentration x dose conversion factor) to a more complex combination of all these model types (Table 3.2).

Each model has five input parameters (a tractable number), each with a specified probability density distribution, similar to the types which often arise in environmental modelling (Table 3.3).

Histograms for the model input distributions based on 500 runs using the two techniques of RS and LHS are shown in Figures 3.11 to 3.15. Output from the two techniques for each of the five models is shown in Figures 3.16 to 3.20. It appears from examination of the figures, that in every case, LHS produces samples from the input distributions that agrees more closely with the expected distributions. Table 3.4 compares the means and variances of the model outputs as calculated using LHS and RS. It is clear from the table that estimates made using the two approaches agree closely.

FIGURE 3·11a

HISTOGRAM OF A (TRIANGULAR DISTRIBUTION) OBTAINED USING RANDOM SAMPLING

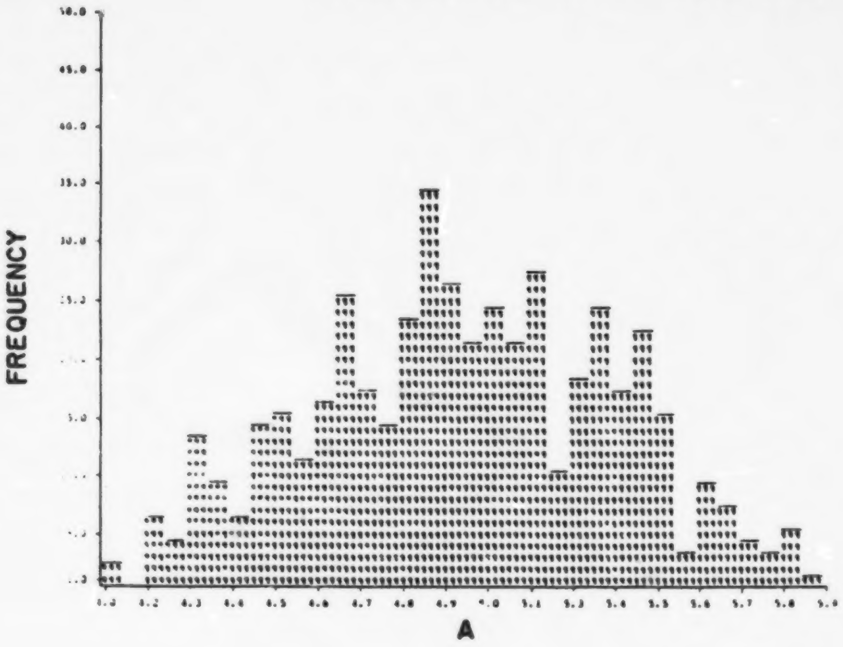


FIGURE 3·11b

HISTOGRAM OF A (TRIANGULAR DISTRIBUTION) OBTAINED USING LATIN HYPERCUBE SAMPLING

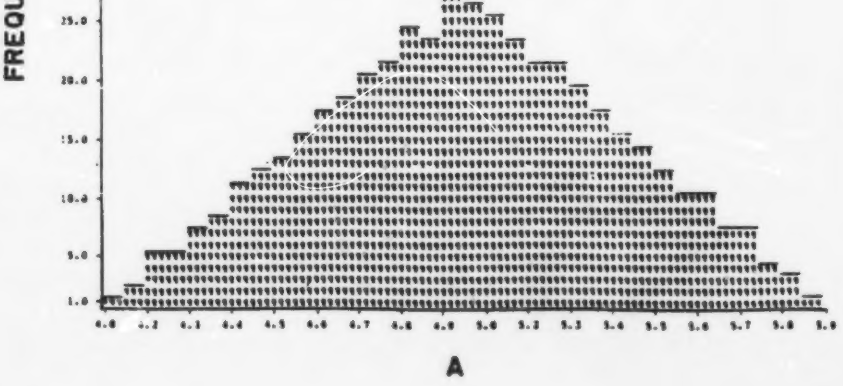


FIGURE 3·12a

HISTOGRAM OF B (UNIFORM DISTRIBUTION)
OBTAINED USING RANDOM SAMPLING

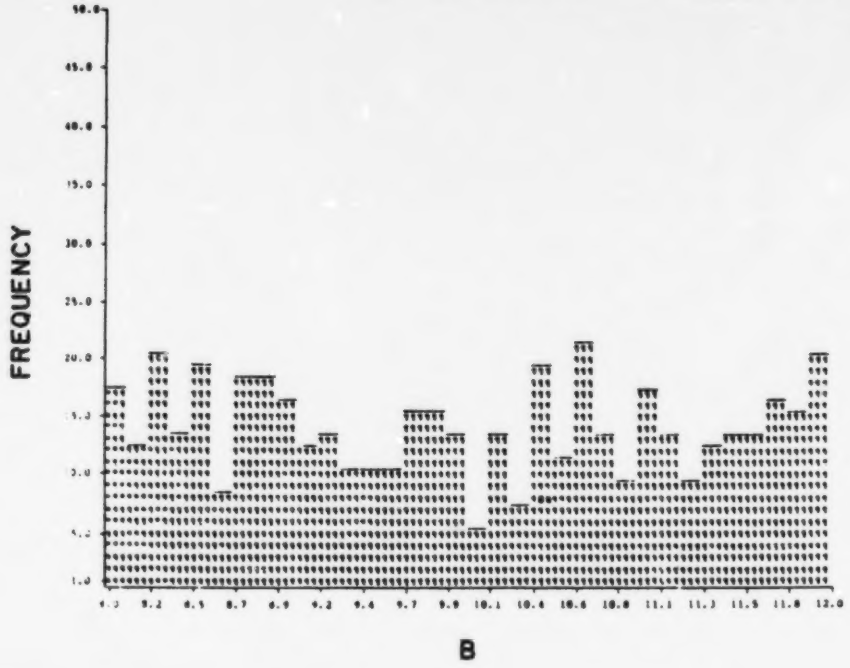


FIGURE 3·12b

HISTOGRAM OF B (UNIFORM DISTRIBUTION)
OBTAINED USING LATIN HYPERCUBE SAMPLING

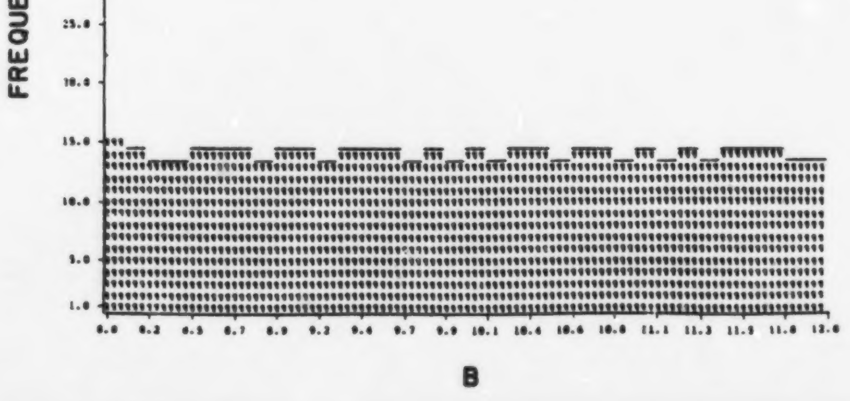


FIGURE 3-13a

HISTOGRAM OF C (LOGUNIFORM DISTRIBUTION)
OBTAINED USING RANDOM SAMPLING

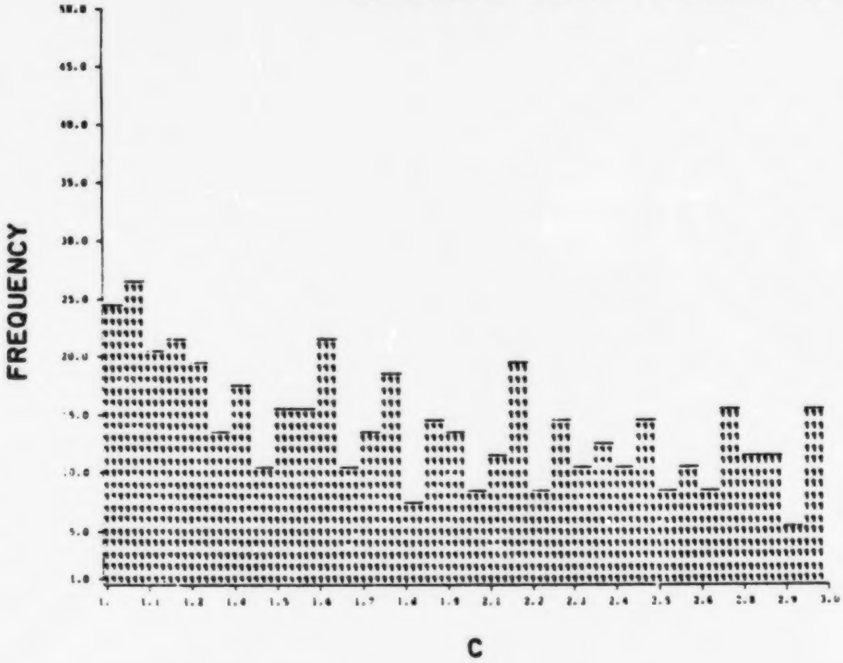


FIGURE 3-13b

HISTOGRAM OF C (LOGUNIFORM DISTRIBUTION)
OBTAINED USING LATIN HYPERCUBE SAMPLING

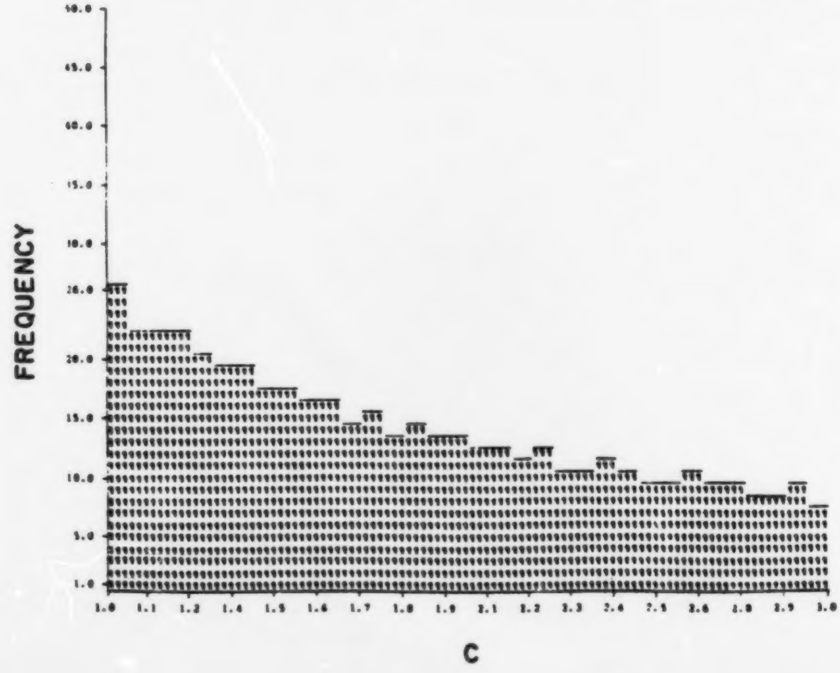


FIGURE 3·14a

HISTOGRAM OF D (LOGNORMAL DISTRIBUTION)
OBTAINED USING RANDOM SAMPLING

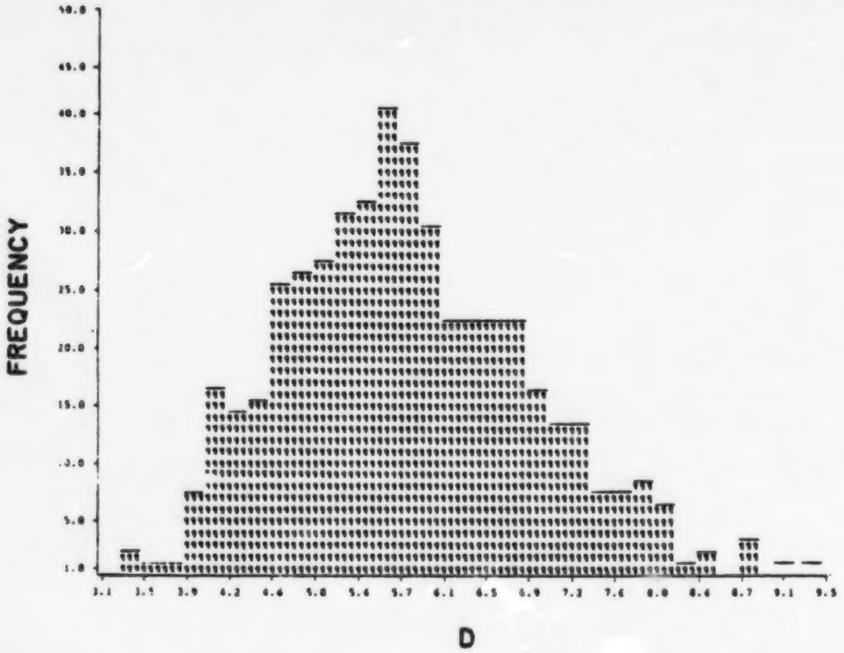


FIGURE 3·14b

HISTOGRAM OF D (LOGNORMAL DISTRIBUTION)
OBTAINED USING LATIN HYPERCUBE SAMPLING

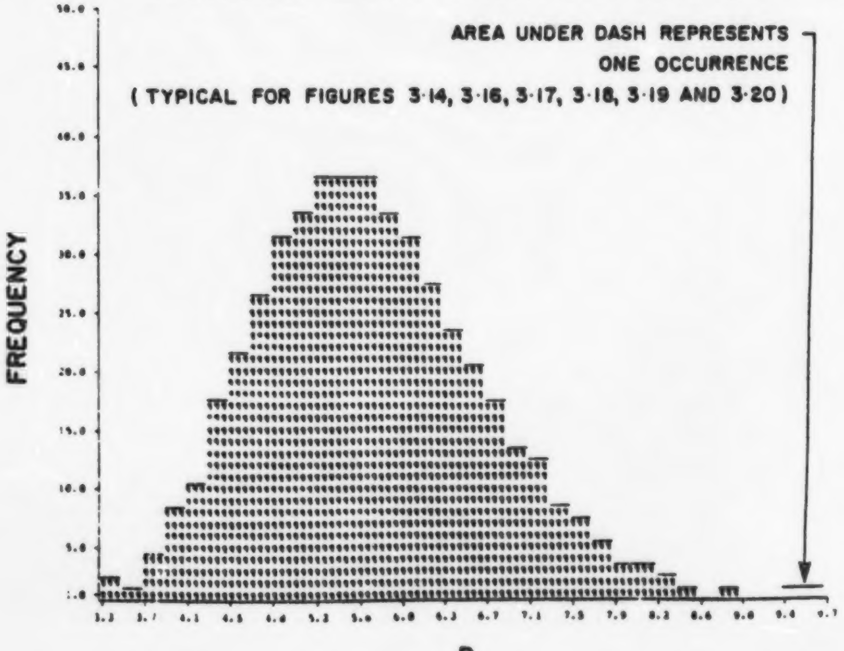

AREA UNDER DASH REPRESENTS
ONE OCCURRENCE
(TYPICAL FOR FIGURES 3·14, 3·16, 3·17, 3·18, 3·19 AND 3·20)

FIGURE 3·15a

HISTOGRAM OF E (NORMAL DISTRIBUTION)
OBTAINED USING RANDOM SAMPLING

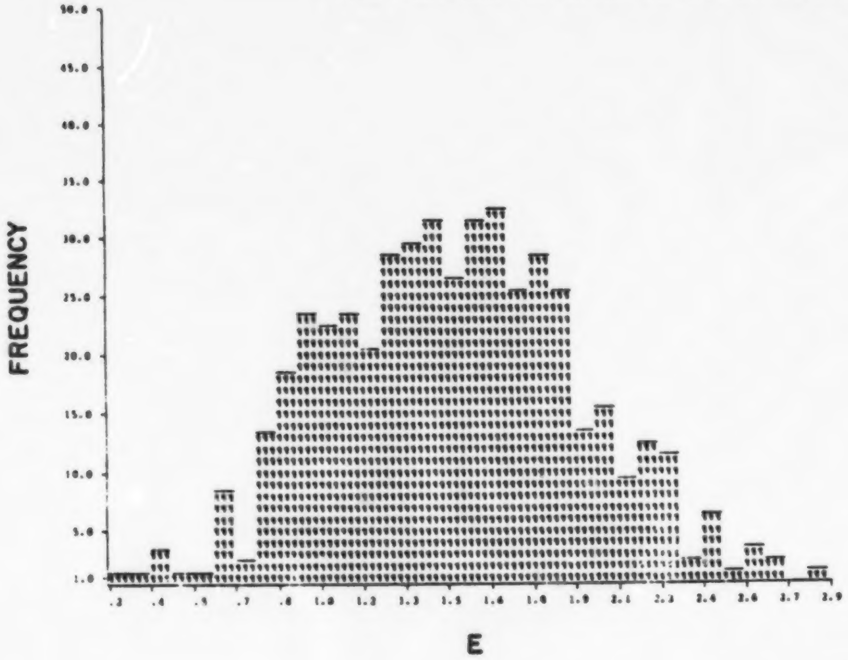


FIGURE 3·15b

HISTOGRAM OF E (NORMAL DISTRIBUTION)
OBTAINED USING LATIN HYPERCUBE SAMPLING

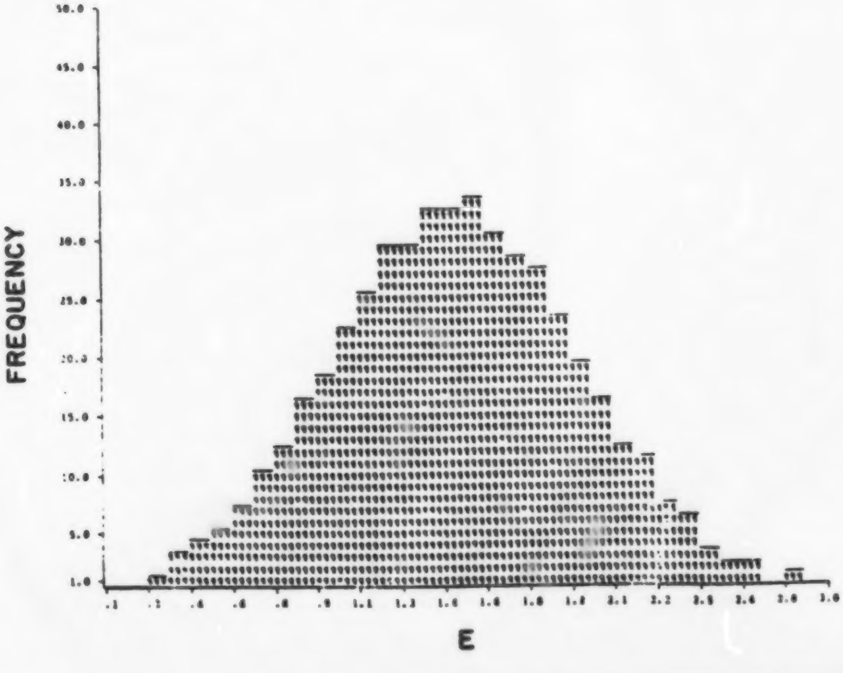


FIGURE 3-16a

HISTOGRAM OF YI
OBTAINED USING RANDOM SAMPLING

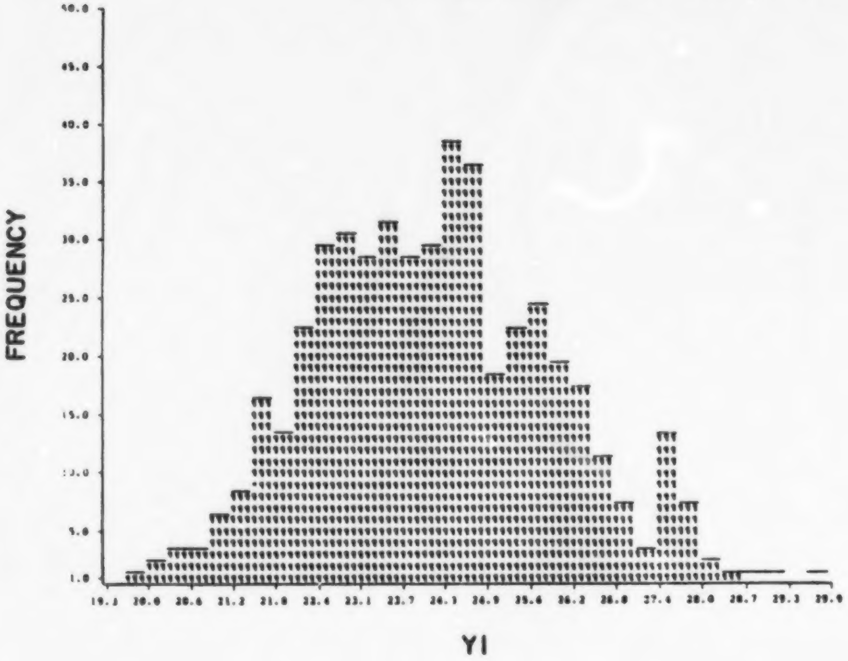


FIGURE 3-16b

HISTOGRAM OF YI
OBTAINED USING LATIN HYPERCUBE SAMPLING

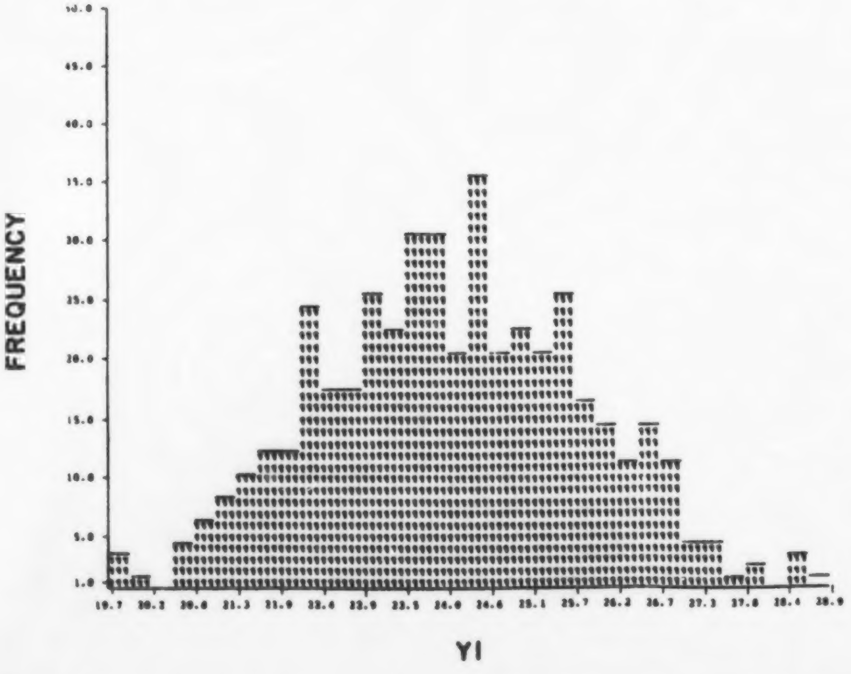


FIGURE 3·17a

HISTOGRAM OF Y2
OBTAINED USING RANDOM SAMPLING

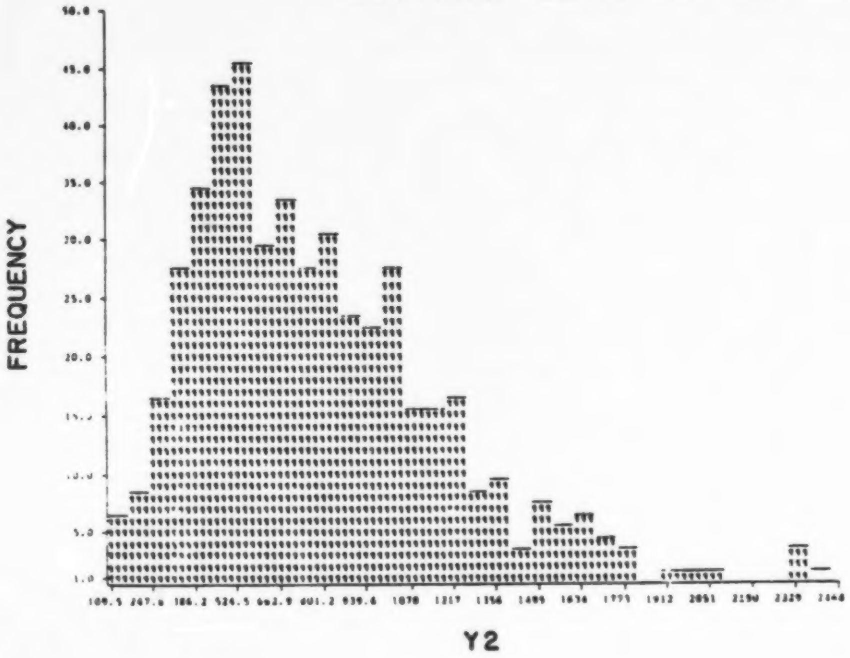


FIGURE 3·17b

HISTOGRAM OF Y2
OBTAINED USING LATIN HYPERCUBE SAMPLING

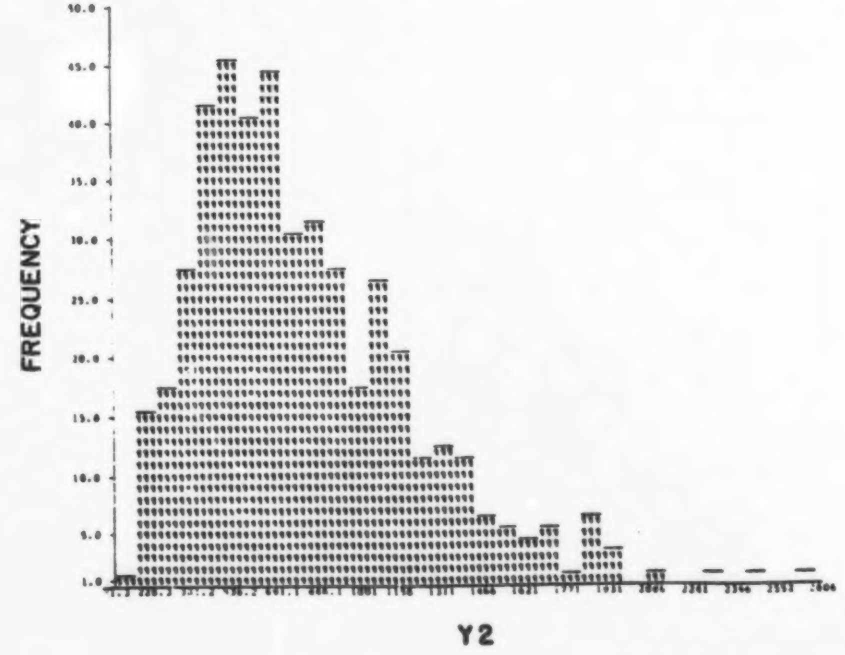


FIGURE 3-18a

HISTOGRAM OF Y3
OBTAINED USING RANDOM SAMPLING

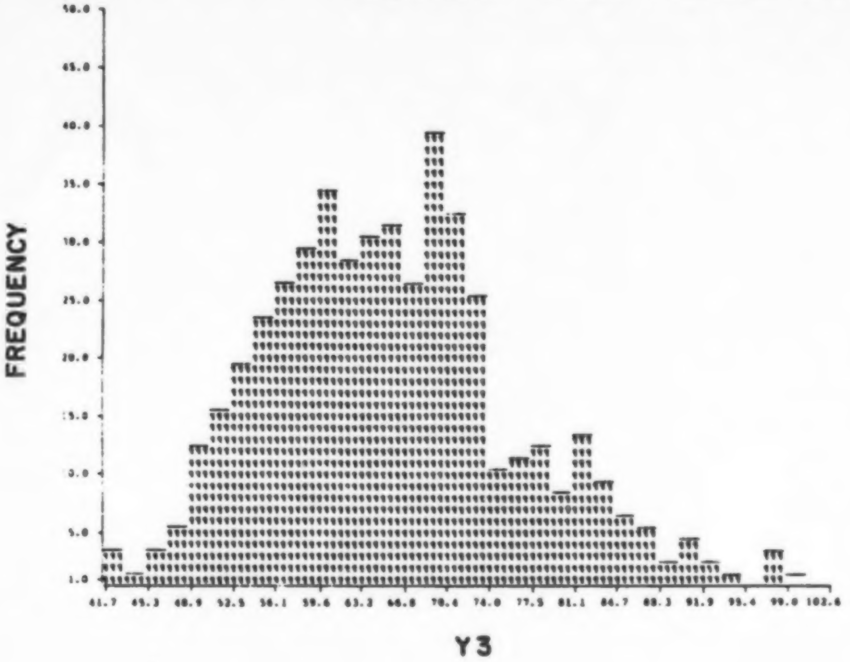


FIGURE 3-18b

HISTOGRAM OF Y3
OBTAINED USING LATIN HYPERCUBE SAMPLING

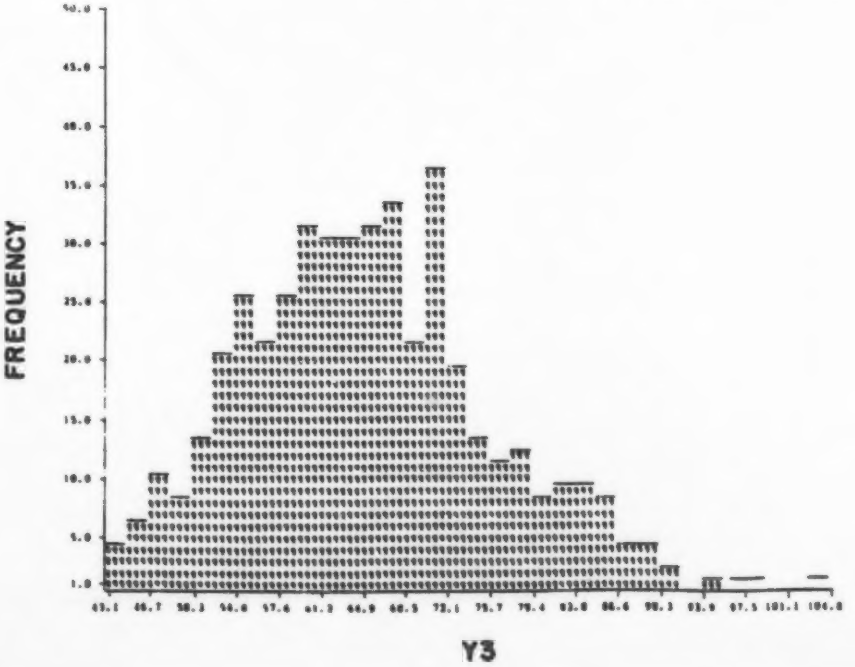


FIGURE 3·19a

HISTOGRAM OF Y4
OBTAINED USING RANDOM SAMPLING

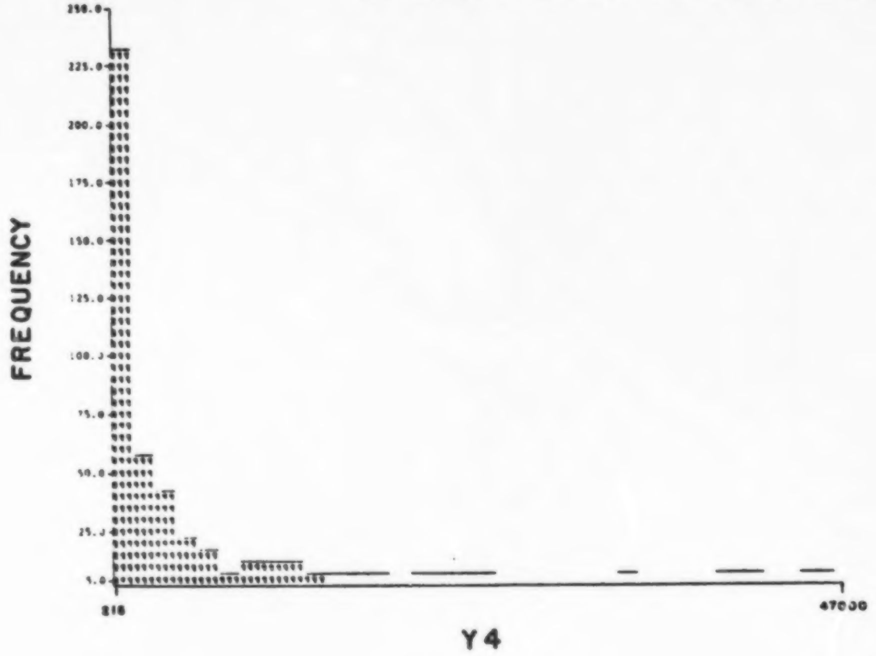


FIGURE 3·19b

HISTOGRAM OF Y4
OBTAINED USING LATIN HYPERCUBE SAMPLING

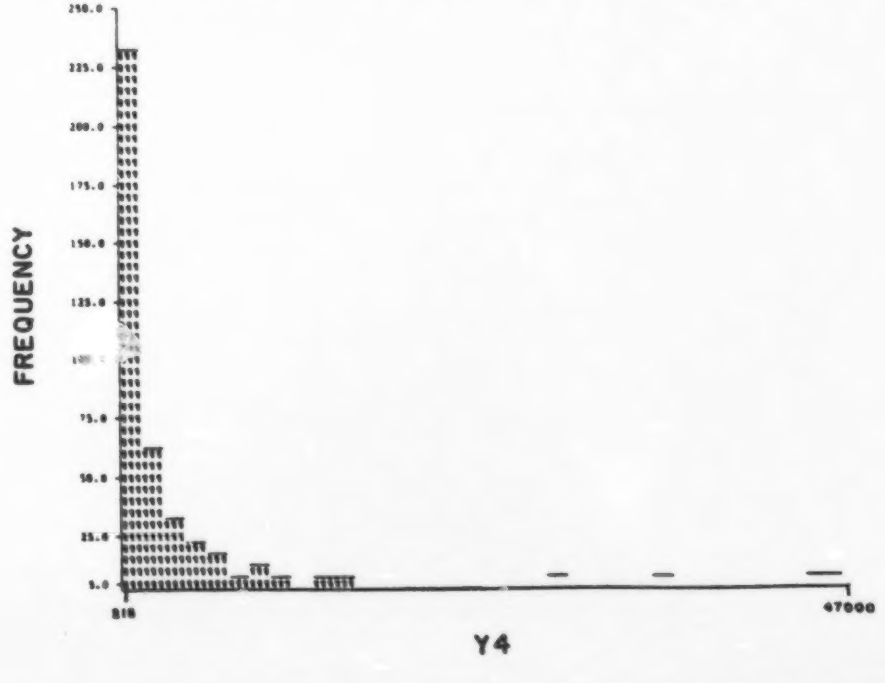


FIGURE 3-20a

HISTOGRAM OF Y5
OBTAINED USING RANDOM SAMPLING

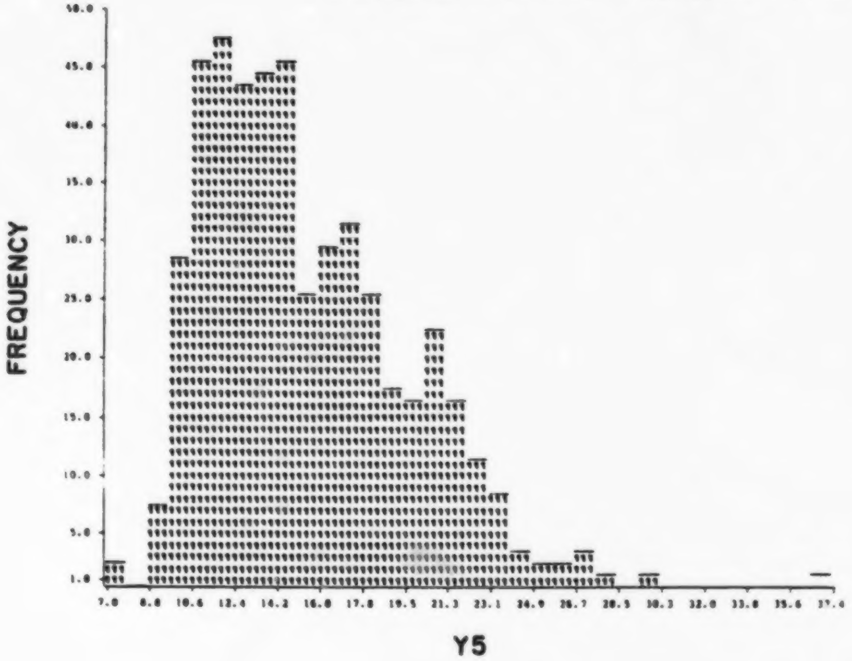


FIGURE 3-20b

HISTOGRAM OF Y5
OBTAINED USING LATIN HYPERCUBE SAMPLING

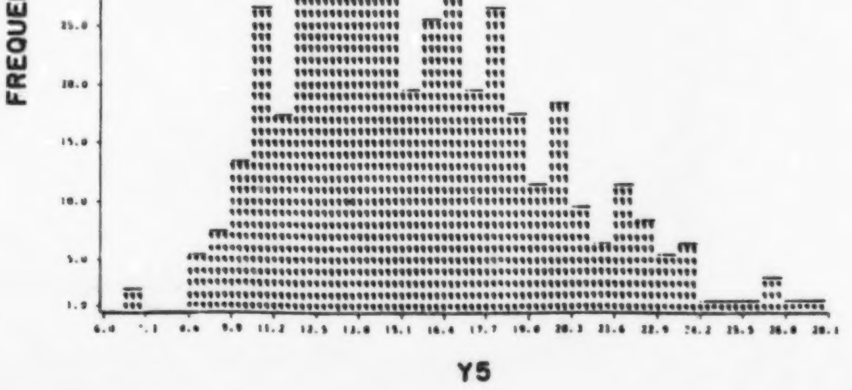


TABLE 3.2

Example Models Used in Comparison of Latin Hypercube and Simple Random Sampling Methods

$$Y1 = A + B + C + D + E$$

$$Y2 = A \times B \times C \times D \times E$$

$$Y3 = (A \times B) + (C \times D \times E)$$

$$Y4 = (A \times B^{C} \times D) + E$$

$$Y5 = A + \frac{B}{C} \times \sqrt{D} \times (1-e^{-E})$$

TABLE 3.3

Input Parameter Distributions Used in Example Models

Parameter	Distribution	Descriptors
A	Triangular	min=4.0 mode=5.0 max=6.0
B	Uniform	min=8.0 max=12.0
C	Loguniform	min=1.0 max=2.0
D	Lognormal	min=3.0 max=10.0 GM=5.66 GSD=1.2
E	Normal	min=0.0 max=3.0 u=1.5 SD=0.5

Notes

min = minimum value
max = maximum value
u = arithmetic mean
SD = arithmetic standard deviation
GM = geometric mean
GSD = geometric standard deviation

TABLE 3.4

Comparison of Means and Variances of Example Models

	LHS[1]		RS[2]	
Model	Mean	Variance[3]	Mean	Variance[3]
Y1	24.1	3.11	24.2	3.33
Y2	792	1.82×10^5	824	1.91×10^5
Y3	65.8	122	66.3	120
Y4	4710	5.35×10^7	5450	6.81×10^7
Y5	15.7	16.8	15.7	19.2

Notes

1. Based on 500 trials of each model using latin hypercube sampling (LHS).
2. Based on 500 trials of each model using simple random sampling (RS).
3. Variance as calculated from the output data.

4.0 SENSITIVITY AND UNCERTAINTY ANALYSIS TECHNIQUES

In probabilistic modelling, the terms sensitivity and uncertainty analysis refer to techniques which seek to define the relationship between model inputs and model predictions. Sensitivity analysis attempts to define the rate of change of model output with respect to changes in model inputs. Uncertainty analysis attempts to relate output variability to the variability in the input parameters. The variability in both input parameters and model predictions is expressed by their probability distribution and associated statistical descriptors. In essence, we are treating the "sample" of probabilistic trials as a series of experiments and we are attempting to isolate the most important parameters measured in these experiments. This chapter examines various techniques for assessing the response of model output results to changes in one or more input parameters and illustrates their use by means of simple numerical examples.

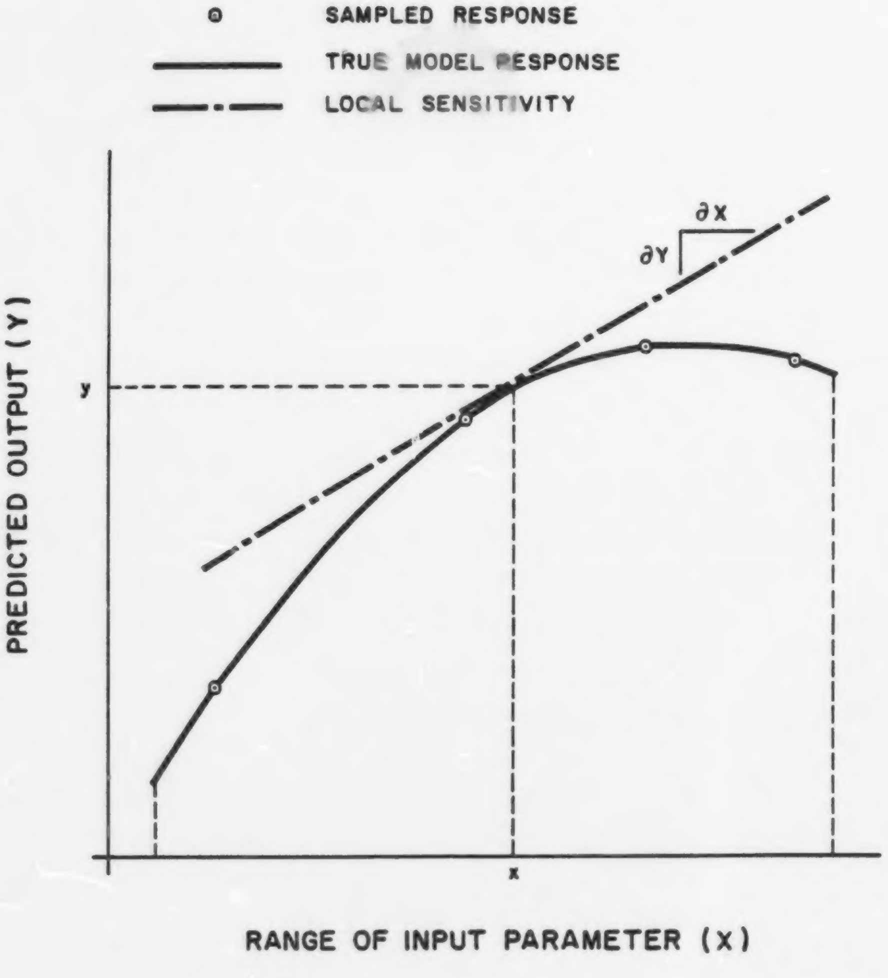
The literature on this subject is extensive. The reader interested in further discussions beyond that provided in this report is referred to recent articles such as those by Downing et al (1985), Gardner (1985), Hoffman and Gardner (1983), Hendrickson (1984), INTERA (1983), and Rose and Swartzman (1981).

4.1 General Discussion of Techniques

Sensitivity analysis is employed to identify those parameters having the greatest influence on the magnitude of model predictions. In the classical sense, sensitivity analysis indicates the rate of change of the model output relative to a change in the various model inputs. Consider for example Figure 4.1. In this case, the classical measure of sensitivity of the output (Y) to the single input parameter(X) is the first order partial derivative, $\delta Y/\delta X$ i.e. the slope of Y(X) at any

FIGURE 4·1

SENSITIVITY OF OUTPUT
TO SINGLE INPUT PARAMETER



particular point x. Such derivatives are typically called "sensitivity coefficients" in the literature.

Uncertainty analysis examines output variability with respect to input variability. The analysis attempts to identify those parameters that contribute most to the uncertainty (i.e. the spread) of the output distribution.

4.1.1 Tomovic Sensitivity and the Slope of the Line

Classical sensitivity methods were first proposed by Tomovic and Vukobratovic (1963) as the rate of change of the state of the system (Y) as a result of a change in a particular parameter (X) (see Figure 4.1). This problem is easily solved for time-invariant parameters and static systems by means of simple calculus (i.e. the partial derivative of Y with respect to X). It is worth noting that the model need not be linear; only the response of the model in a suitably small region around the nominal value of the parameters needs to be linear.

The primary uses of Tomovic sensitivity analysis are to explore responses to small changes in inputs, initial conditions, and parameters; reveal functional relationships within the model; and test the local stability of model behaviour near some equilibrium point (typically the mid-point of the region). Tomovic sensitivity has been widely used in mathematical and engineering applications.

The problems with Tomovic sensitivity, at least for the purposes of environmental assessment, are: the equations, although conceptually simple, can be extremely difficult to develop and numerical solutions (i.e., computer perturbation experiments) are often required; and Tomovic sensitivity measures the linear responses of the model to small changes in individual parameters. Thus, Tomovic sensitivity is less useful when parameters are simultaneously varied over large ranges, and can give inadequate and misleading information in these situations (Gardner et al 1981).

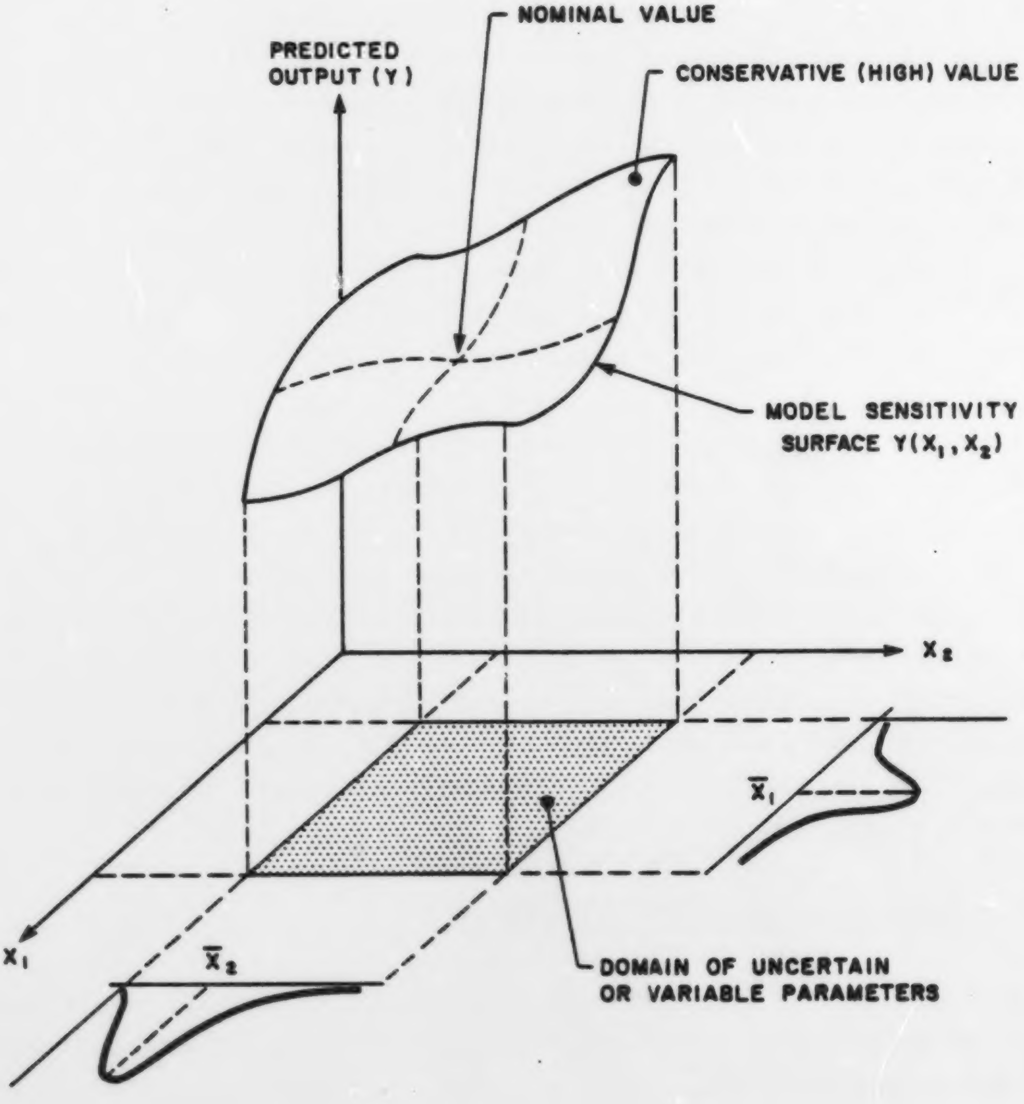
To illustrate the foregoing, consider Figure 4.1. In this instance, it is assumed that the output (Y) is determined by a single input parameter (X). This is not a real restriction, since if Y depends on more than one input parameter, the partial derivative of the output variable Y with respect to the parameters of interest would be calculated and the remaining parameters would be held constant, most often at their average value. It is clear that the sensitivity of Y (i.e. $\delta Y/\delta X$) depends on where the derivative is evaluated. Thus Tomovic (or classical) sensitivity provides only a local measure of sensitivity.

The situation is more complex when the output variable depends on more than one parameter. Consider for example the situation illustrated in Figure 4.2. In this case, the output Y depends on two parameters X_1 and X_2. The sensitivity of Y with regard to X_1 and X_2 is obtained by calculating the partial derivatives $\delta Y/\delta X_1$ and $\delta Y/\delta X_2$ at a particular point on the surface determined by the possible combinations of X_1 and X_2 (i.e. the "domain" of X_1 and X_2). However, for the case illustrated by Figure 4.2, the response surface is "bumpy" and again the partial derivative provides only a local measure of sensitivity. If the surface were planar or nearly planar, the partial derivative would provide a global measure of sensitivity.

In instances where the partial derivatives can be calculated with relative ease, the derivatives can be evaluated at several values of X_1 and X_2 to provide an indication of how non-linear the response surface actually is. (This is related to the

FIGURE 4·2

ILLUSTRATION OF SENSITIVITY AND UNCERTAINTY OF OUTPUT TO TWO INPUT PARAMETERS


NOMINAL VALUE
PREDICTED OUTPUT (Y)
CONSERVATIVE (HIGH) VALUE
MODEL SENSITIVITY SURFACE $Y(X_1, X_2)$
X_2
$\overline{X}_1$
X_1
$\overline{X}_2$
DOMAIN OF UNCERTAIN OR VARIABLE PARAMETERS

uncertainty in the linear regression coefficient b_i as discussed in the next section.) Typically, the partial derivatives would be evaluated at nominal parameter values (e.g. average values) to obtain a single, overall estimate of sensitivity.

As a matter of interest, in conventional deterministic analysis (i.e. using single estimates of parameter values), it is usual to choose parameter values which lead to high output values (i.e. conservative values). Such a point is shown on Figure 4.2. It is clear that the deterministic value is not representative of the output; rather, it represents the output value produced by an unlikely combination of values of X_1 and X_2. A somewhat more representative estimate of Y would be obtained if parameter values closer to the centre of the distributions (e.g. average values) were used in the evaluation.

Numerical techniques can be used to estimate the differential sensitivities of a model. For example, Gardner (1984) has observed that if all parameters are varied simultaneously around their nominal values with variances set at 1% of the nominal value, then the partial derivative of model response to a given parameter can be estimated by least squares methods. Gardner notes that this technique may encounter numerical problems if the variances are smaller than 1%; while if the variances are (typically) larger than 1%, interactions between parameters that are correlated may affect the results.

4.1.2 Regression Analysis Techniques

The advent of fast computers and the development of Monte Carlo simulation methods made possible the statistical characterization of model predictions. As discussed in Chapter 3, Monte Carlo methods involve the iterative selection of random values of model parameters from specified frequency distributions. A distribution of model predictions is thus obtained that is a function of the random set of parameters. Uncertainties and sensitivities are then derived from the Monte Carlo simulations by analyzing the combined set of parameters (X_i's) and predictions (Y's).

In some instances, response surface techniques are used to replace a complex mathematical simulation model with a simpler model that still produces similar results. Sensitive and uncertain parameters can then be identified from an examination of the predictions from the simpler response surface model. (See for example, Downing et al 1985).

The data set produced from Monte Carlo simulations allows the of many different measures of dependencies of the Y's on the X_i's. Methods based on the theory and rigor of correlation and regression procedures are particularly suited to, and have been extensively used, to examine the sensitivity of the output prediction (Y) with respect to the input parameters (X_i's) (via estimation of the partial derivatives). The methods also provide a means for identifying those inputs which contribute most to the output uncertainties. A full discussion of regression analysis can be found in Draper and Smith (1981) and Edwards (1979).

In the following discussion, a linear regression model of the following form is used:

$$Y = a + b_1X_1 + b_2X_2 + \ldots\ldots b_nX_n \qquad (4\text{-}1)$$

where Y (the prediction) is the dependent variable, the X_i's (inputs) are the independent variables and the b_i's are the conventional regression coefficients (RC's) determined by the regression analysis. The model represented in equation (4-1) is both linear in the regression parameters (i.e. a, b's) and first order in the independent variables (the X_i's). This form of regression model was used since the regression coefficients (i.e. the b's) and various associated statistical quantities provide a simple mean for identifying those regression variables which contribute most to the sensitivity and uncertainty of the output values (the Y's).

The method of least squares is widely used to derive the regression coefficients. The total sum of squares for the Y variable can be partitioned into two components: total sum of squares = regression sum of square + residual sum of squares or

$$\sum_{i} (Y_i - \overline{Y})^2 = \sum_{i}^{n} (Y_p - \overline{Y})^2 + \sum_{i}^{n} (Y_p - Y_i)^2 \qquad (4\text{-}2)$$

where Y_p = predicted value for Y_i from the regression equation (4-1)

Y_i = observed Y values from simulation model

$\overline{Y}$ = estimated mean value of Y from simulation model

n = number of observed values

In essence, the objective of a regression analysis is to replace the actual response surface of output Y (see Figure 4.2) with a surface which is closely approximated by the Y's calculated from equation (4-1). The regression equation is generally much simpler and easier to handle than the complex probabilistic assessment model used to generate the output Y. The larger the portion of variability in the output which is accounted for by the regression, the better the regression is able to reflect the actual response surface. If the regression model provides a good approximation of the behavior of the "real" model (as indicated by goodness-of-fit criteria), reliable conclusions can be made relating to sensitivity and uncertainty from quantities generated in the regression. Considerable caution must, however, be exercised if the fitted model is used with different inputs than those used to develop it. The existence of singularities in the original model also confounds the analysis.

Various statistical indicators can be used to describe the adequacy of the linear regression model as a predictor of the assessment model's behaviour and the contribution of different input variables to output uncertainty. Some indicators of interest in sensitivity and uncertainty analysis are discussed below.

The multiple correlation coefficient R^2 provides an indication of the ability of the regression model to predict the behaviour of the dependent variable. The overall R^2 is merely 1 minus the ratio of the residual sum of squares of the regression model to the total sum of squares or

$$R^2 = 1 - \left[\sum (Y_p - Y_i)^2 / \sum (\bar{Y} - Y_i)^2 \right] \qquad (4\text{-}3)$$

Clearly the better the model predicts Y, the smaller the predictive error and the closer R^2 is to 1.

A measure of the confidence one has in the prediction of the regression coefficients (i.e. the b_i's of equation (4-1)) is provided by the value of the overall F statistic. The F statistic is calculated as the mean sum of squares due to regression divided by the mean sum of squares of the residual error. These quantities are derived by dividing the sum of squares quantities in equation (4-2) by the appropriate degrees of freedom (Draper and Smith (1981)).

The regression coefficients (RC's) from a multiple regression analysis of the model output (Y) on the input parameters (the

X_i's) provide an indication of the rate of change in the output variable Y with respect to a unit change in each input variable, namely $\delta Y/\delta X_i$ (Gardner 1984). The RC's are equivalent to the partial derivatives evaluated at the mean value of the input parameters. The RC's provide useful information on the influence of the mean value of the input parameters on the location of the (mean) value of the output.

The computed F value for each regression coefficient (not to be confused with the overall F statistic noted earlier) provides a good indicator of how significant a particular parameter is to the regression. The computed F value for the ith regression coefficient (F_i) is calculated as follows:

$$F_i=b_i{}^2/SE_i{}^2 \tag{4-4}$$

where

b_i = regression coefficient for the ith parameter

SE_i = standard error of the regression coefficient for the ith parameter. SE_i is the square root of the total residual mean square error, an estimate of the variance about the regression, divided by the total variance attributable to variable X_i. Clearly the larger the variance of X_i the smaller the standard error of b_i. A smaller standard error in the denominator of equation (4-4) will lead to a larger (and hence more significant) F statistic.

As indicated by equation (4-4), F_i takes account of both slope (a measure of sensitivity) and uncertainty (via $SE_i{}^2$) and is thus a combined measure of sensitivity and uncertainty. The F_i values for the regression coefficients on the respective parameters (X_i's) can be compared with tabular listings of F values for the appropriate degrees of freedom to determine the probability of a particular regression coefficient not being zero (i.e. of being significant). The larger the F_i for a particular X_i, the more significant that particular variable is in the regression.

The F value for a regression coefficient can also be decomposed into a semi-partial correlation coefficient r_{sp} (see for example Edwards 1979) as follows:

$$r_{sp_i} = F_i \ (1-R^2)/(n-k-1) \qquad (4\text{-}5)$$

where

n = sample size
k = number of independent variables

According to Edwards (1979), the following statement can be made about the ith independent variable (X_i) with respect to the ith correlation coefficient: "Then we know that when X_i is entered last in the regression equation, it accounts for r_{spi} of the total sum of squares over and above that accounted for by the other independent variables." This procedure would seem to provide an excellent indication of the contribution of an input parameter to output uncertainty in that it performs the same "experiment" (namely being entered last in the regression) on each parameter. Since r_{sp} for each variable is merely the F statistic for the regression coefficient (i.e. F_i) multiplied by a constant, examination of the relative magnitudes of the F statistic provides a good relative measure of a contribution of the parameter to the uncertainty in the model output. Some examples to illustrate the use of the F statistic for the regression coefficients are discussed in Section 4.2.

As indicated in the foregoing, linear regression methods provide a useful technique for examining the relationships between a dependent variable and one or more independent variables. A measure of the adequacy of the linear regression model is provided by R^2. The closer R^2 is to 1, the more adequate is the regression model. In many instances, the R^2 value can be improved by a transformation of variables, such as a log-transformation (see Section 4.2).

Rank transform can also be used in regression analysis to linearize the data. In rank transform, the data are replaced with their corresponding ranks and the ranks entered into the regression model. Iman and Conover (1979) suggest that the rank transform method is a simple, repeatable process that compares favourably with other methods. For further discussion of this technique, the interested reader is referred to the extensive literature (e.g. Helton and Iman 1982; Conover 1980).

One further note is appropriate. In situations where it is not possible to designate at least one variable as being independent, then correlation is the appropriate method for describing the relationship between two (or more) mutually dependent variables. The coefficient of correlation describes the degree of association between the variable. On the other hand, the regression coefficient described previously measures the size of the change in the dependent variable which can be predicted when a unit change is made in an independent variable. The reader interested in further discussion of the relation between regression coefficients and correlation coefficients is referred to the literature (e.g. Draper and Smith 1981; Conover 1980).

4.1.3 Sensitivity Index

The sensitivity index SI (Hoffman and Gardner 1983) provides an indication of sensitivity and uncertainty provided output values are always positive. The model is run with the ith parameter set at its high and low limits, while all other parameters are held constant at their nominal values. The SI_i is calculated as follows:

$$SI_i = 1 - \frac{\text{low output result for the ith parameter}}{\text{high output result for the ith parameter}} \qquad (4-6)$$

The closer the SI is to one (i.e. the larger the difference between the high and low results), the more sensitive the model is to variable X_i.

4.1.4 Graphical Techniques

Various graphical approaches have been used in sensitivity analysis. Typically, one parameter is allowed to vary while all other parameters are held constant at their nominal values. The model output is then plotted against the variable of interest. If the limits of the parameter of interest are indicated on the graph, the sensitivity index can also be calculated. Non-linearities and rates of change of the output versus a particular input variable are also indicated by this technique. It must be noted that since all other parameters are held constant, this technique only accounts for the slope at a particular point on the response surface.

A graphical approach for combined sensitivity and uncertainty analysis has been outlined by Wuschke et al (1983) of Atomic Energy of Canada Limited (Whiteshell Nuclear Research Establishment). In this approach, the parameter of interest is held constant at a chosen value in its range while all other parameters are allowed to vary within their specified distributions. Typically, the parameter under examination is held constant at the maximum and minimum values of its distribution. The results of the high and low runs are compared with those of a reference case (in which all parameters are varied) by graphically superimposing the cumulative plots. Selected statistics of the output results can also be compared. This provides a comparison between effects of different levels of variables given the range of variation that exists in the remaining variables.

4.1.5 Uncertainty Index

The variance of the result of a perfectly additive or multiplicative model can be estimated by combining the variances of the input parameters. Assuming statistical independence between the X_i's, for a purely additive model of the form

$$Y_a = X_1 + X_2 + X_3 .. + X_n \tag{4-7}$$

the variance of Y_a is simply the sum of the variances of the individual parameters:

$$\mathrm{Var}(Y_a) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) \ldots + \mathrm{Var}(X_n) \tag{4-8}$$

Similarily, a purely multiplicative model of the form

$$Y_m = X_1 * X_2 * X_3 .. * X_n \tag{4-9}$$

can be changed into an additive model by a log transformation:

$$\ln Y_m = \ln X_1 + \ln X_2 + \ln X_3 + \ldots + \ln X_n \tag{4-10}$$

The variance of $\ln Y_m$ is then equal to:

$$\mathrm{Var}(\ln Y_m) = \mathrm{Var}(\ln X_1) + \mathrm{Var}(\ln X_2) + \mathrm{Var}(\ln X_3) + \ldots + \mathrm{Var}(\ln X_n) \tag{4-11}$$

which is readily determined from the geometric standard deviation (GSD):

$$\mathrm{Var}\ (\ln Y_m) = [\ln\ (\mathrm{GSD}(Y_m))]^2 \qquad (4\text{-}12)$$

Since variance is a measure of parameter uncertainty, the ratio of the variance of a particular parameter to the overall variance of the model result provides an estimate of the contribution of that parameter to model uncertainty. A ratio using this concept has been termed the "Importance Index" (Hoffman and Gardner 1983). However, preference has been given in this study to the use of the term "uncertainty index" to maintain consistency with the earlier definition of uncertainty analysis (see also SENES 1984).

For an additive model (equation 4-7), the uncertainty index (UI) is given by:

$$UI_i = \frac{\mathrm{Var}\ (X_i)}{\mathrm{Var}\ (Y_a)} \qquad (4\text{-}13)$$

For a multiplicative model (equation 4-9), the uncertainty index is calculated by:

$$UI_i = \frac{\mathrm{Var}\ (\ln X_i)}{\mathrm{Var}\ (\ln Y_m)} = \frac{[\ln(\mathrm{GSD}(X_i))]^2}{[\ln(\mathrm{GSD}(Y_m))]^2} \qquad (4\text{-}14)$$

Unfortunately, this index is only strictly valid for purely additive or multiplicative models. For such models, the parameter uncertainty indices would all be less than 1.0 and would also sum to 1.0. A parameter uncertainty index near 1.0 would indicate that the particular parameter contributed significantly to uncertainty in the result. Conversely, an index near zero indicates little contribution to total variation in the results and may suggest that the parameter could be assigned a suitable "constant" value in subsequent analyses with little loss of information.

4.1.6 Other Techniques

In addition to the techniques discussed already in this section, a wide variety of other techniques for sensitivity and uncertainty analysis have been reported in the literature. including the adjoint method of sensitivity analysis (see for example, Oblow 1983).

In a recent paper on the use of response-surface methodologies for uncertainty analysis in assessment models, Downing et al (1985) reviewed several techniques, including the adjoint method, for screening the results of an uncertainty analysis to identify which inputs should receive the most attention. These authors concluded that while the adjoint method is theoretically appealing, it is impractical, even with only moderate size computer codes, due to the difficulty in obtaining the adjoint solutions.

As discussed in Section 4.3, our view is that regression analysis is the single most useful tool for sensitivity and uncertainty analysis. However, no single technique is ever likely to always be "best" and as many techniques as possible, within the constraints of a particular study, should be used.

4.2 Some Examples Used to Illustrate Methods of Sensitivity and Uncertainty Analysis

A series of generic mathematical models were developed and run probabilistically to illustrate some of the sensitivity/ uncertainty analysis techniques discussed in the previous section. Four of the generic models introduced in Section 3.5.9 were selected to illustrate the essential features of sensitivity and uncertainty analysis (see Table 4.1). Several different probabilistic cases were run for the models using the input distributions shown in Table 4.2. As discussed earlier, although using purely mathematical models might seem abstract, it allows the reader to readily examine the mathematical interactions in the model, the input distributions and the relationship between these factors and the output variability.

The sensitivity and uncertainty indicators presented in the following section need to be examined carefully. Although the discussions are detailed (necessarily), the reader should look critically at the formulae and the inputs and ask if the information conveyed by the indicators is consistent with the reader's prior expectations. This is a similar process to sensitivity/uncertainty analysis of "real" models except that, in the real case, the reader goes one step further and asks if the sensitivity/uncertainty information is consistent with what he knows about the physical system. In essence, the sensitivity/uncertainty analysis is a method of critically examining the model with regard to the relationship between its inputs and outputs.

4.2.1 Model Y1

Let us first examine the sensitivity/uncertainty indicators shown in Table 4.3 for model Y1. The regression coefficients (RC's) and partial derivatives are all 1.0; hence, they were not included in the table. Since the regression model is identical to the "real" model (both are purely additive models), the b's are all 1.0 and indicate that a unit change in any parameter will produce the same change in Y1 for all possible values of the input parameter. Similarly, because the regression model exactly

Table 4.1

Example Models Used to Illustrate Techniques of Sensitivity/Uncertainty Analysis

$$Y1 = A + B + C + D + E$$

$$Y2 = A \times B \times C \times D \times E$$

$$Y3 = (A \times B) + (C \times D \times E)$$

$$Y5 = A + \frac{B}{C} \times \sqrt{D} \times [1-e^{-E}]$$

Table 4.2

Input Parameter Distributions Used in Example Models

CASE I A, B, C, D, E uniformly distributed from 1 to 3, u = 2

CASE II A uniformly distributed from 1 to 10, u = 5.5;
B, C, D, E uniformly distributed from 1 to 3, u=2

CASE III A uniformly distributed from 0 to 4, u = 2;
B, C, D, E uniformly distributed from 1 to 3, u = 2

CASE IV A triangular distribution (0,1,5), u = 2;
B, C, D, E uniformly distributed from 1 to 3, u = 2

Case V Distributions as indicated below.

Parameter	Distribution	Descriptors
A	Triangular	min=4.0 mode=5.0 max=6.0
B	Uniform	min=8.0 max=12.0
C	Loguniform	min=1.0 max=2.0
D	Lognormal	min=3.0 max=10.0 GM=5.66 GSD=1.2
E	Normal	min=0.0 max=3.0 u = 1.5 SD= 0.5

Notes

min = minimum value
max = maximum value
u = arithmetic mean
SD = arithmetic standard deviation
GM = geometric mean
GSD = geometric standard deviation

Table 4.3

Sensitivity/Uncertainty Indices for Model Y1[4] for Five Cases

Parameter	A		B		C		D		E	
Case[1]	SI[2]	UI[3]	SI	UI	SI	UI	SI	UI	SI	UI
CASE I	0.18	0.20	0.18	0.20	0.18	0.20	0.18	0.20	0.18	0.20
CASE II	0.50	0.87	0.14	0.05	0.14	0.05	0.14	0.05	0.14	0.05
CASE III	0.33	0.51	0.18	0.13	0.18	0.13	0.18	0.13	0.18	0.13
CASE IV	0.31	0.40	0.18	0.15	0.18	0.15	0.18	0.15	0.18	0.15
CASE V	0.08	0.04	0.15	0.44	0.08	0.11	0.25	0.38	0.12	0.06

Notes

1. See Table 4.2 for description of cases.

2. $SI = 1 - \frac{\text{low output result}}{\text{high output result}}$

3. $UI = \frac{\text{VAR (input)}}{\text{VAR (output)}} = \frac{\ln (\text{GSD})^2 \text{ (input)}}{\ln (\text{GSD})^2 \text{ (output)}}$

4. $Y1 = A + B + C + D + E$

predicts behaviour of the "real" model, the F statistic for all regression coefficients is infinitely large; thus, the F statistics are not included in Table 4.3. However, since the model is purely additive, the uncertainty indices (UI) are a good uncertainty indicator and sum to exactly 1.0. Note that the sensitivity index changes for each run and is hence not a "case independent" sensitivity index.

4.2.2 Model Y2

The pure multiplicative model, Y2, when log-transformed, behaves in precisely the same way as model Y1 in a regression analysis. One can, by examining the form of a mathematical model, make preliminary judgements as to what type of regression (untransformed, log-transformed, etc.) will provide a good fit. (The higher the R^2 value derived from the regression, the larger is the fraction of variation about the mean explained by the regression and the better is the fit.) Because adding logarithms is the same as multiplying untransformed numbers, the log-transformed regression model exactly predicts the real model.

Results of the sensitivity/uncertainty analysis on models Y2 are summarized in Table 4.4. With a purely multiplicative model, the UI behaves similarily to the additive model and is a good uncertainty indicator. Calculated UI's, using geometric standard deviations, sum to a value of 1.0 as before. In cases III, IV and V, the lowest value of one of the inputs is zero. This causes the sensitivity index to equal 1 for the purely multiplicative model. In this situation, the sensitivity index loses its mathematical robustness and is not a reliable indicator.

Table 4.4

Sensitivity/Uncertainty Indices For Model Y2[4] for Five Cases

Parameter	A		B		D		D		E	
Case [1]	SI[2]	UI[3]	SI	UI	SI	UI	SI	UI	SI	UI
Case I	0.67	0.20	0.67	0.20	0.67	0.20	0.67	0.20	0.67	0.20
Case II	0.90	0.44	0.33	0.14	0.33	0.14	0.33	0.14	0.33	0.14
Case III	1.0	0.71	0.33	0.08	0.33	0.08	0.33	0.08	0.33	0.08
Case IV	1.0	0.52	0.66	0.13	0.66	0.13	0.66	0.13	0.66	0.13
Case V	0.33	0.02	0.33	0.04	0.50	0.40	0.70	0.12	1.0	0.46

1. See Table 4.2 for description of cases

2. $SI = 1 - \frac{\text{low output result}}{\text{high output result}}$

3. $UI = \frac{\text{Var (input)}}{\text{Var (output)}} = \frac{\ln (GSD)^2 \text{ (input)}}{\ln (GSD)^2 \text{ (output)}}$

4. $Y2 = A \times B \times C \times D \times E$

4.2.3 Model Y3

The sensitivity/uncertainty indicators for model Y3 are shown in Tables 4.5 and 4.6. Model Y3 is a simple hybrid model with both additive and multiplicative terms. Note how closely the regression coefficient (RC) predicts the partial derivative (PD) (Table 4.5).

Consider first the partial derivatives and the regression coefficients for case I (A,B,C,D,E uniformly distributed from 1 to 3). They tell us that the model is twice as sensitive to parameters C,D, and E (RC ≅ 4.0, PD = 4.0) as it is to parameters A and B (RC ≅ 2.0, PD = 2.0). This can be understood, for say A and C, as follows. By holding B constant and varying A to its limits, AxB ranges from 2 to 6 or a difference of (6-2=)4 units in the output Y3. By holding D and E constant and varying C to its limits, CxDxE ranges from 4 to 12, or a difference of (12-4=)8 units in the output Y3. This indicates that a change in C (or D or E) can produce twice the change in Y3 as the same change in A (or B).

Since in case I the inputs are identical, the F statistic should give us similar information to the RC and indeed it does. Since model Y3 is neither purely additive nor purely multiplicative, the UI's do not sum to one; however, at least in the cases shown for Y3, they still give a relative indication of the contribution of a particular parameter to model uncertainty.

In case II, A has been changed to a uniform distribution ranging from 1 to 10. Note that although A has been changed, the PD and RC for A have not changed from case I; instead, the PD and RC for B have changed. The F statistic indicates that A is of primary importance in the regression, with B second and C,D, and E jointly third. The function of the F statistic as a combined

Table 4.5

Partial Derivatives and Regression Information for Model $Y3^5$ for Five Cases

Parameter	A			B			C			D			E			
Case[1]	PD[2]	RC[3]	F[4]	PD	RC	F	PD	RC	F	PD	RC	F	PD	RC	F	R^{2*}
CASE I	2.0	2.0	354	2.0	2.1	421	4.0	4.2	1632	4.0	4.2	1653	4.0	4.1	1624	0.92
CASE II	2.0	2.0	3096	5.5	5.8	1406	4.0	4.3	763	4.0	4.2	762	4.0	4.0	685	0.92
CASE III	2.0	2.0	1146	2.0	2.2	365	4.0	4.3	1370	4.0	4.2	1381	4.0	4.1	1323	0.92
CASE IV	2.0	2.0	1031	2.0	2.2	371	4.0	4.2	1416	4.0	4.2	1419	4.0	4.1	1381	0.94
CASE V	10.0	9.5	1487	5.0	4.9	3618	8.6	8.6	2899	3.0	2.8	964	11.5	11.5	2888	0.96

Notes

1. See Table 4.2 for description of cases
2. Partial Derivative
3. Regression Coefficient
4. F statistic
5. Y3 = (A x B) + (C x D x E)

* R^2= multiple correlation coefficient from regression analysis

Table 4.6

Sensitivity/Uncertainty Indices for Model Y3[4] for Five Cases

Parameter	A		B		C		D		E	
Case[1]	SI[2]	UI[3]	SI	UI	SI	UI	SI	UI	SI	UI
CASE I	0.29	0.01	0.29	0.01	0.50	0.01	0.50	0.01	0.50	0.01
CASE II	0.36	0.12	0.45	0.01	0.34	0.01	0.34	0.01	0.34	0.01
CASE III	0.50	0.05	0.29	0.01	0.50	0.01	0.50	0.01	0.50	0.01
CASE IV	0.56	0.04	0.29	0.01	0.50	0.01	0.50	0.01	0.50	0.01
CASE V	0.26	<0.01	0.26	0.01	0.23	<0.01	0.25	0.01	0.11	<0.01

Notes

1. See Table 4.2 for description of cases.
2. $SI = 1 - \frac{\text{low output result}}{\text{high output result}}$
3. $UI = \frac{\text{Var (input)}}{\text{Var (output)}} = \frac{\ln (\text{GSD})^2 \text{ input}}{\ln (\text{GSD})^2 \text{ output}}$
4. $Y3 = (A \times B) + (C \times D \times E)$

sensitivity/uncertainty indicator is well illustrated here with variable B's contribution going up from case I to case II with absolutely no change in the input distribution for B. The UI's, although not summing to 1, are still useful in case II as indicators of the relative contribution of each input variable to output uncertainty.

On first examination, it appeared somewhat surprising that the RC's and PD's for B changed from case I to case II due only to a change in A. The next two cases were used to examine further this aspect (Table 4.5). In case III, A was changed to a uniform distribution with a range from 0 to 4 and a mean value of 2. In case IV, A was changed to a triangular distribution with a range from 0 to 5 a peak at 1 and a mean value of 2. In both cases III and IV, B, C, D, and E were uniformly distributed between 1 and 3. Cases III and IV preserve the same mean value of 2 for all input distributions that were used in case I. Note that all of the RC's are the same as case I! In case V, when all of the distributions are changed, the RC's bear no resemblance to the previous cases (as is to be expected since the partial derivatives are dependent on both the form of the model and the values of the independent variable at the point at which the derivatives are evaluated).

It must be concluded that none of the sensitivity indicators discussed in this report is "input independent"; that is, they depend on both model form and input values (indeed, we are not aware of any input-dependent indicator). The RC's provide sensitivity information for a specific set of inputs. This information cannot necessarily be translated to another case with a new set of inputs. With this particular model (Y3), the RC's remained the same only when the mean values of the input distributions remained unchanged. (This was also found to hold for all five models.) Thus, any sensitivity analysis of "real" models must be done carefully.

4.2.4 Model Y5

Model Y5 (see Tables 4.7 and 4.8 for sensitivity/uncertainty indices) is complex in structure and is representative of many "real" environmental models. It also provides an excellent illustration of some of the potential pitfalls of uncertainty analysis. If the reader examines cases I to IV, he will immediately notice that the RC's and F statistics behave in a completely consistent manner. The RC's remain constant in all four cases and the F statistics go up when the variance of an input variable is increased. Does this mean that the RC's are a pure "input independent" sensitivity indicator? No, it does not! This can be seen by examining case V and case VI, in which the RC's change. Case VI, is a special additional case run for model Y5. The nature of case VI is described at the bottom of Tables 4.7 and 4.8. It provides valuable insight as to why the RC's change.

The change in the PD's and RC's for C,D, and E can be explained by examining the partial derivatives of model Y5, namely:

$$\frac{\delta Y5}{\delta A} = 1.0$$

$$\frac{\delta Y5}{\delta B} = \frac{\sqrt{D}}{C} \times (1-e^{-E})$$

$$\frac{\delta Y5}{\delta C} = -\frac{B\sqrt{D}}{C^2} \times (1-e^{-E})$$

$$\frac{\delta Y5}{\delta D} = \frac{B}{2C \times \sqrt{D}} \times (1-e^{-E})$$

$$\frac{\delta Y5}{\delta E} = \frac{B\sqrt{D}}{C} \times (e^{-E})$$

Table 4.7

Partial Derivatives and Regression Information for Model Y5[5] for Five Cases

Parameter	A			B			C			D			E			
Case[1]	PD[2]	RC[3]	F[4]	PD	RC	F	PD	RC	F	PD	RC	F	PD	RC	F	R^{2*}
CASE I	1.0	1.0	4356	0.60	0.64	1931	0.60	-0.69	2220	0.30	0.32	468	0.19	0.23	255	0.95
CASE II	1.0	1.0	88135	0.60	0.64	1931	0.60	-0.69	2220	0.30	0.32	468	0.19	0.23	255	0.99
CASE III	1.0	1.0	17414	0.60	0.64	1931	0.60	-0.69	2220	0.30	0.32	468	0.19	0.23	255	0.98
CASE IV	1.0	1.0	14280	0.60	0.64	1930	0.60	-0.69	2219	0.30	0.32	468	0.19	0.23	255	0.97
CASE V	1.0	0.93	34	1.00	1.1	419	5.60	-5.8	2883	0.89	0.88	232	2.9	3.5	729	0.99
CASE VI**	1.0	0.99	298	0.60	0.65	2790	1.67	-1.9	1160	0.83	0.86	244	0.52	0.66	145	0.96

Notes

1. See Table 4.2 for description of cases
2. Partial Derivative
3. Regression Coefficient
4. F statistic
5. $Y5 = A + \frac{B}{C} \times D \times (1-e^{-E})$

* R^2 = multiple correlation coefficient from regression analysis

** CASE VI: A,C,D,E = uniform from 1 to 3
B = uniform from 1 to 10

Table 4.8

Sensitivity/Uncertainty Indices for Model Y5[5] for Five Cases

Parameter	A		B		C		D		E	
Case[1]	SI[2]	UI[3]	SI	UI	SI	UI	SI	UI	SI	UI
CASE I	0.48	0.45	0.32	0.45	0.18	0.45	0.37	0.45	0.14	0.45
CASE II	0.89	0.92	0.18	0.05	0.22	0.05	0.10	0.05	0.07	0.05
CASE III	0.76	0.74	0.32	0.19	0.18	0.19	0.37	0.22	0.14	0.22
CASE IV	0.80	0.72	0.32	0.22	0.18	0.22	0.37	0.22	0.14	0.19
CASE V	0.12	<0.01	0.23	0.07	0.52	0.02	0.33	0.06	0.12	0.01
CASE VI[4]	0.33	0.06	0.68	1.27	0.51	0.06	0.28	0.06	0.22	0.06

Notes

1. See Table 4.2 for description of cases.
2. $SI = 1 - \frac{\text{low output result}}{\text{high output result}}$
3. $UI = \frac{\text{Var (input)}}{\text{Var (output)}} = \frac{\ln (GSD)^2 \text{ (input)}}{\ln (GSD)^2 \text{ (output)}}$
4. CASE VI: A,C,D,E = uniform from 1 to 3
   B = uniform from 1 to 10
5. $Y5 = A + \frac{B}{C} \times \sqrt{D} \times [1-e^{-E}]$

In the present analysis, the partial derivatives were evaluated at the mean values of the input parameters (see Section 4.1.1). Since B is the only variable whose mean value has changed in case VI, it follows that the value of partial derivatives dependent on B will change; in this case, only A is independent of B.

Further investigation of case V, which involved widening the geometric standard deviation (GSD) of the lognormally distributed variable D, revealed some interesting trends. As the GSD of D was widened, the R^2 of the untransformed regression decreased and the RC's became steadily poorer predictors of the partial derivatives. When Y5 and A,B,C,D and E were logtransformed and used in the multiple regression, the R^2 increased and so did its success in predicting the partial derivative (once the RC's from the logtransformed regression were converted back to untransformed values). This indicates, not too surprisingly, that the higher the R^2 of a regression, the better will be the sensitivity/uncertainty information that is provides.

The UI's, in the first four cases provide a good indication of the relative contribution to uncertainty by each input parameter. However, in case V, the output variance is large relative to any of the input variances and the UI as an indicator of uncertainty begins to break down.

4.3 Discussion

The cases discussed in this chapter are by no means meant to provide an exhaustive explanation of sensitivity/uncertainty analysis. Rather, they are meant to illustrate the application of sensitivity/uncertainty analysis tools to specific problems and to encourage the reader to look critically at these tools.

Even in the simple examples presented in this section, complexities arise. Sensitivity/uncertainty analysis is somewhat of a art. It is not a rote, automatic means of identifying sensitive and uncertain variables, but rather is part of a feedback loop between the sensitivity/uncertainty analysis, the model and the real system being modelled. It should stimulate questions that must be rationally answered rather than answering questions in an uncritical fashion.

In terms of which sensitivity/uncertainty analysis tools are best, no single method of analysis is ever likely to be always best and as many tools as time permits should be applied.

Tomovic (classical) sensitivity is useful during the development of a model and for initial screening of parameter importances. For instance, if a parameter has a high Tomovic sensitivity index and a large variance, then this parameter will contribute significantly to model uncertainties. If, on the other hand, the parameter has a low sensitivity value combined with a relatively small variance, then it is unlikly to make a significant contribution to the variance of model predictions. Problems arise when parameters have a low Tomovic sensitivity value and high variance, or vice versa. In these cases, Tomovic sensitivity is of limited use (Hoffman and Gardner 1984).

Many techniques are available for numerical analysis of model dynamics. Because most assessments are performed with less than adequate information, simple measures of dependency between predictions and input parameters are often the most useful and reliable. In this regard, if only a single tool is to be applied the authors recommend thoughtful and judicious use of regression analysis. The regression coefficients and F statistics obtained from regression analysis, with a significant multiple correlation coefficient, are particularly useful. Estimates of the sensitivity coefficients (i.e. the partial derivatives) provide some insight into the effect that each parameter may have on model uncertainty. For instance, parameters with high sensitivity and high variance will dominate model uncertainties. Likewise, parameters with low sensitivities and small variances will have little effect on model uncertainties. The behaviour of parameters with intermediate sensitivities and variances will be difficult to predict. For this latter group, the second order effects and interactions among these parameters (values not estimated by the linear sensitivity coefficients) may dominate model uncertainties.

Regression analysis (if a good fit can be obtained) provides maximum information with minimum expenditure of time and, if used with caution, will yield good results.

5.0 CONSIDERATIONS IN QUANTIFYING STATISTICAL CONFIDENCE IN PROBABILISTIC MODELLING RESULTS

Previous chapters of this report have discussed various types and sources of uncertainty in assessment modelling, techniques for propagating uncertainty in input parameters through the assessment model, and techniques for evaluating how sensitive the uncertain output from such an assessment is to a change in one or more input parameters. This chapter discusses some ways to express confidence concerning various attributes of the distribution of output from a probabilistic assessment. First, however, it is appropriate to review some useful features of probabilistic modelling.

5.1 Some Concepts of Probabilistic Modelling

The results (output) from a probabilistic assessment can be regarded as a sample of results drawn randomly from the population of all possible results. Simply put, given the necessary input distributions, a random sampling (i.e. Monte Carlo) procedure is used to find values of the output distribution.

The calculation procedure is repetitive. At each repetition, a single observation is made on each input variable according to its probability distribution. These values of the input variables are substituted into the mathematical model to obtain a single value for each output variable. In essence, this can be regarded as a single "experiment". Then when several repetitions have been made, the set of numerical values for an output variable may be considered as a sample on an objective random variable corresponding to that output variable. From the sample, the mean, variance, or probability function may be estimated as desired. This information may be interpreted as describing the output variable in terms of subjective probability which tells us in quantitative terms what we are entitled to believe about the output variable.

At this point a distinction should be made between uncertainty inherent in the value of the output variable which arises from uncertainty in the input variables and uncertainty about the output variables arising from the sampling error associated with the Monte Carlo process. The latter aspect is the subject of this chapter. The former aspect is communicated quantitatively by the probability density function for the output parameter and remains uncertain in the real world regardless of how much the sampling error is reduced. Observations made in the real world provide a means of testing or validating model predictions (see also Chapters 2 and 6).

The output from a probabilistic analysis is often presented in the form of sample frequency diagrams (histograms) and cumulative frequency diagrams. The sample frequency curves are not identical to the population probability curves but they are expected to resemble the population probability curves more closely as the sample size (i.e. the number of trials or repetitions) increases.

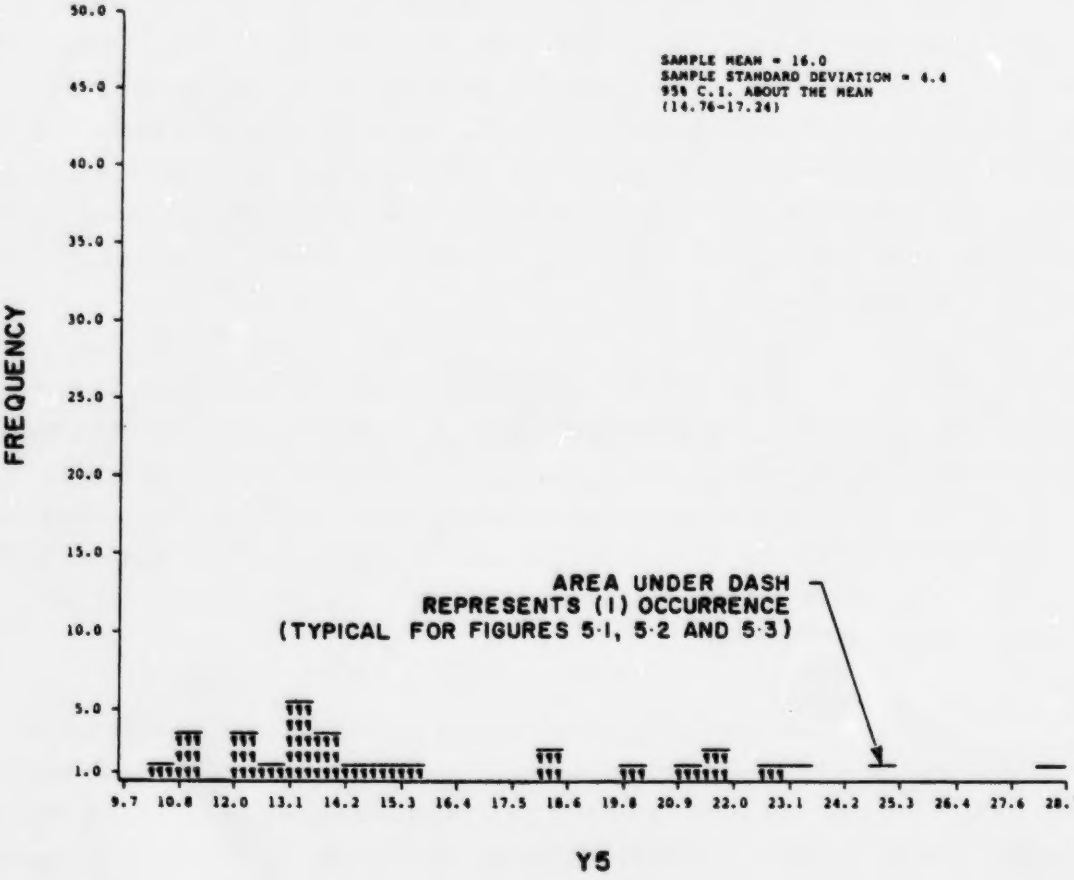
The frequency distribution shown on Figure 5.1 is a result of 50 trials using simple random sampling to choose input parameter values. These values were then substituted in model Y5 to calculate 50 output values. The distribution is not very smooth as a result of the sampling errors that occur in small samples. However, even with 50 trials, there is a suggestion that the distribution may be slightly skewed with the mode to the left of the graph and a long tail extending to the right of the graph.

Figures 5.2, 5.3 and 5.4 show the frequency distributions

FIGURE 5·1

HISTOGRAM OF Y5
WITH CASE V INPUTS AND 50 ITERATIONS

(SEE TABLE 4·2 FOR INPUTS)



SAMPLE MEAN = 16.0
SAMPLE STANDARD DEVIATION = 4.4
95% C.I. ABOUT THE MEAN
(14.76-17.24)

AREA UNDER DASH
REPRESENTS (1) OCCURRENCE
(TYPICAL FOR FIGURES 5·1, 5·2 AND 5·3)

FIGURE 5·2

HISTOGRAM OF Y5
WITH CASE V INPUTS AND 100 ITERATIONS
(SEE TABLE 4·2 FOR INPUTS)

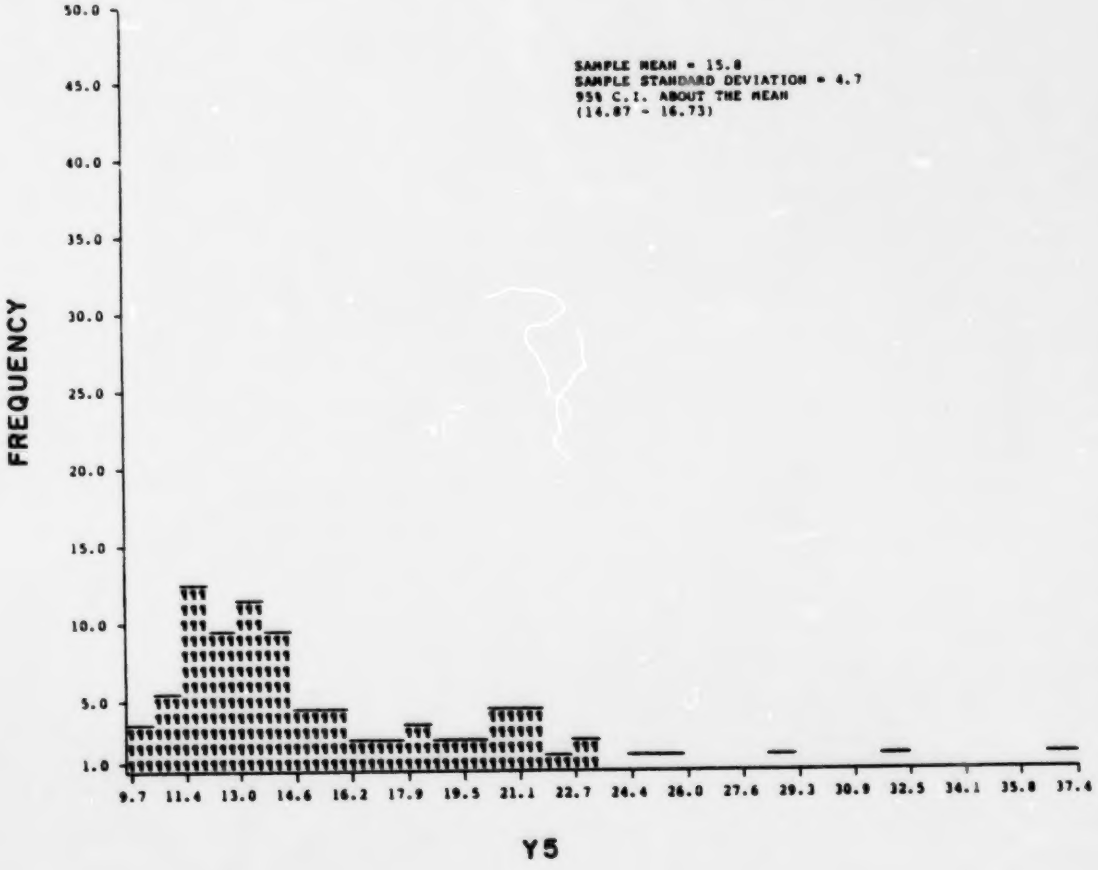


FIGURE 5·3

HISTOGRAM OF Y5
WITH CASE V INPUTS AND 200 ITERATIONS
(SEE TABLE 4·2 FOR INPUTS)

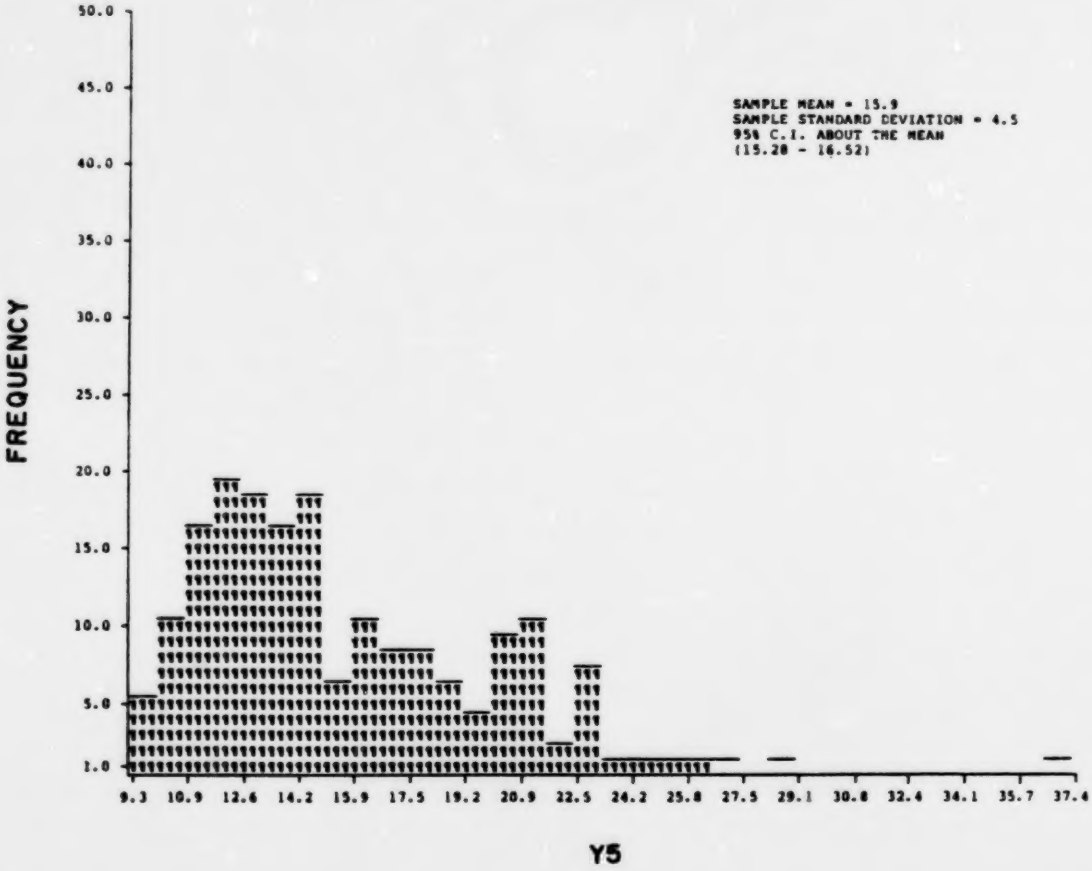
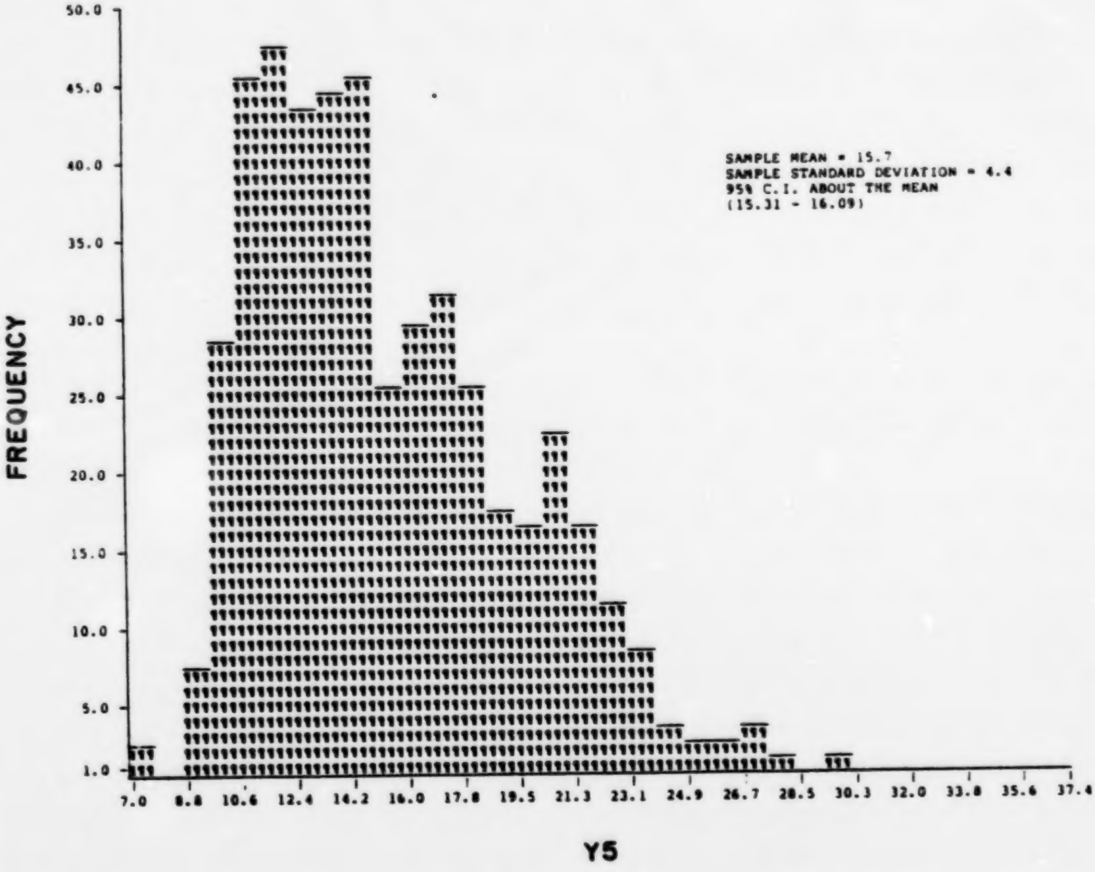


FIGURE 5·4

HISTOGRAM OF Y5
WITH CASE V INPUTS AND 500 ITERATIONS
(SEE TABLE 4·2 FOR INPUTS)



resulting from 100, 200 and 500 Monte Carlo trials of the same model and input parameter distributions used to generate the distribution shown on Figure 5.1. As the number of trials increases, the output frequency distribution (actually the relative frequency distribution) becomes increasingly smoother and approaches more closely the underlying population distribution (which would require an infinite number of trials to define perfectly).

A different aspect of increasing the number of trials is also indicated by the data shown on Figures 5.1 and 5.4. Namely there is very little change in the mean value as the number of trials (n) increases from 50 to 500. The 95% confidence interval on the mean (also indicated on Figures 5.1 to 5.4) narrows in proportion to $1/\sqrt{n}$ as expected. However, as noted previously, many trials are necessary to obtain a reasonable representation of the output curve itself. Thus, it is clear that the number of trials required in any particular situation depends to a large degree on the desired end-use of the evaluation.

Figure 5.5 shows the cumulative frequency plot of the same data shown on Figure 5.4. (Actually, the cumulative relative frequency curve is a step function. However, in the situation shown for 500 trials, the steps are very small and a smooth curve is shown on the Figure 5.5.) Each point on the graph represents the cumulative frequency that the value will be less than or equal to any specified value. In Figure 5.6, the complementary cumulative frequency curve (i.e. downward cumulative frequency distribution) is shown for the same data illustrated in Figures 5.4 and 5.5. A point on Figure 5.6 represents the cumulative frequency that any particular value will be exceeded.

FIGURE 5·5

CUMULATIVE RELATIVE FREQUENCY
(500 TRIALS OF MODEL Y5 WITH MIXED PARAMETERS)

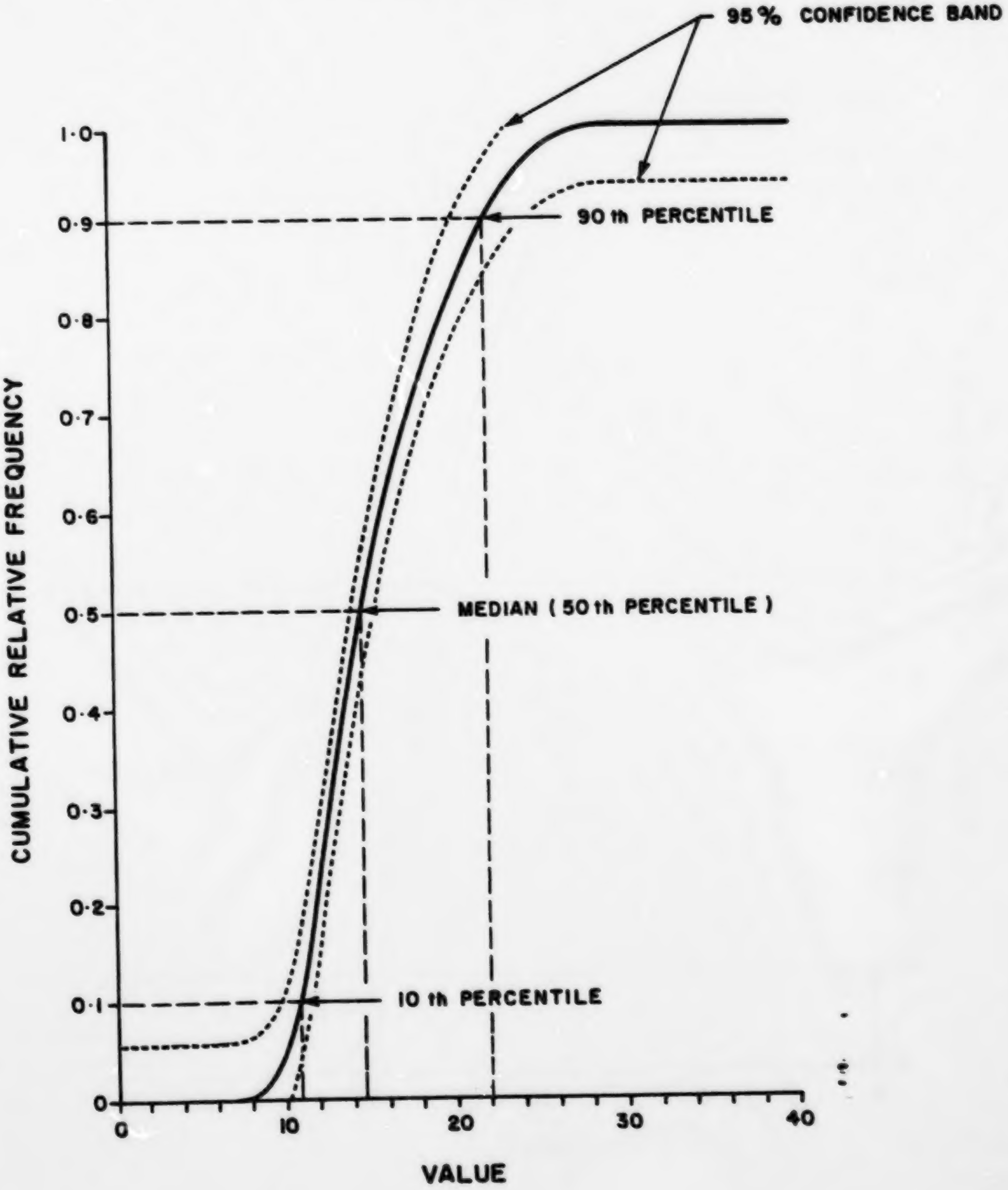
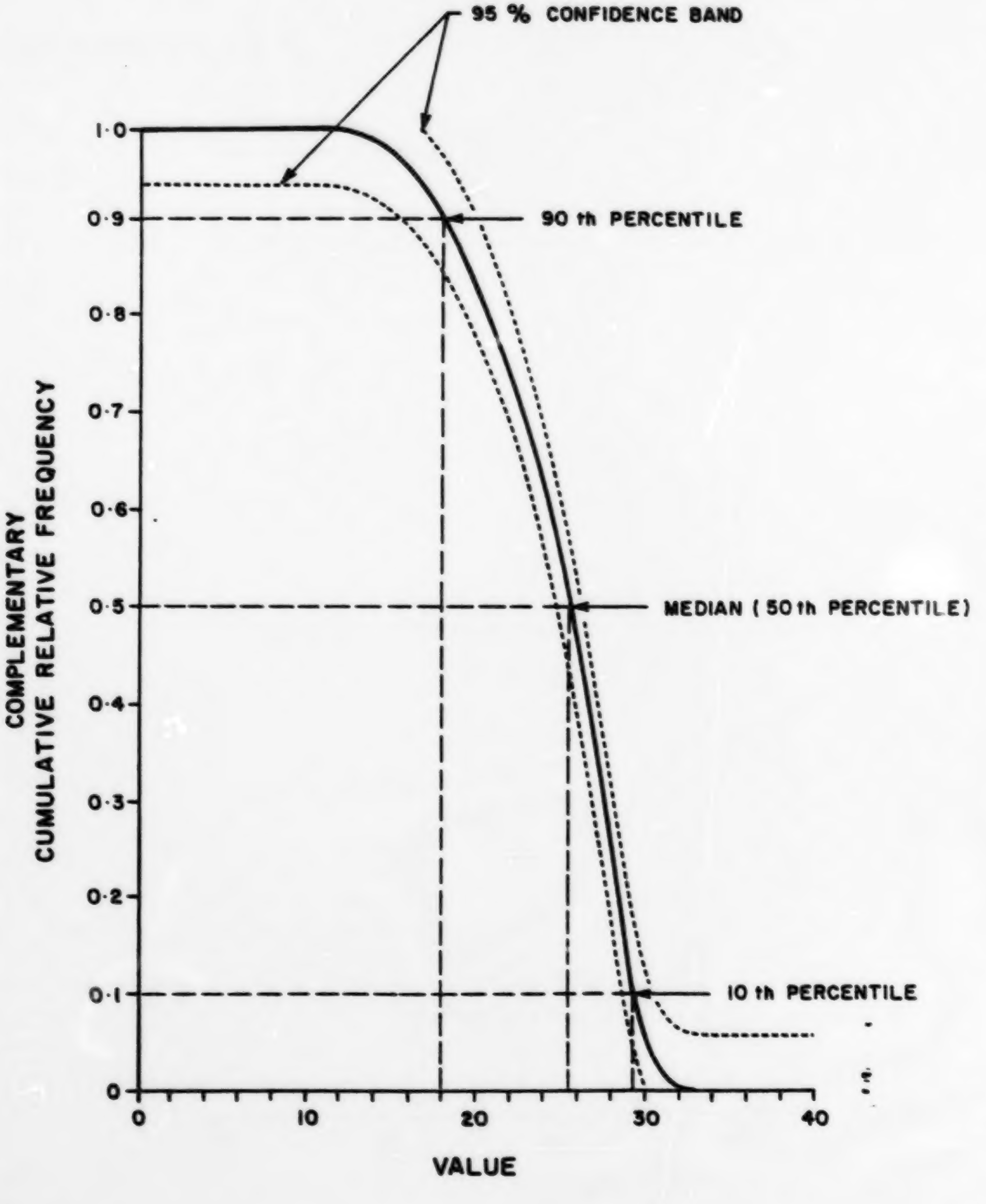


FIGURE 5·6

COMPLEMENTARY CUMULATIVE RELATIVE FREQUENCY (500 TRIALS OF MODEL Y5 WITH MIXED PARAMETERS)



Frequency (or relative frequency) histograms such as those shown in Figures 5.1 to 5.4 or the cumulative frequency diagrams (Figures 5.5 and 5.6) provide a good picture of the sample as a whole. For example, a steeply rising cumulative frequency distribution (such as that shown in Figure 5.5) indicates that most of the values are clustered around a central point, such as mean or median. On the other hand, a slowly rising cumulative relative frequency distribution would indicate larger scatter (dispersion) and hence a relatively larger uncertainty (or variability) in the data.

A distribution can be characterized through various statistics including arithmetic and geometric mean and variance and various percentiles. However, it is important to remember that any attempt to summarize the information presented in graphs such as Figures 5.1 to 5.6 results in a loss of information. For example, if only the mean and variance are reported, then information regarding the extreme values of the results is lost. Whether or not such information is meaningful is a separate issue; the point here is simply that information is lost. If only extreme values are summarized (eg. maximum and minimum values), one is left with no idea of what the most probable value might be.

5.2 Interval Estimates

Sample statistics (such as the mean and variance) provide estimates of the corresponding parameters of the underlying population. The assumption that samples are randomly drawn from the underlying population is important to ensure that the samples are representative of the underlying population and that in turn sample statistics are fair estimates of the population parameters. However, sample statistics such as sample mean and variance, while providing estimates of the corresponding population parameters are not, in general, exactly correct because of sampling variation. Thus, it is important to infer from the properties of the sample the range within which the (population) parameter is likely to lie.

Confidence intervals form one type of interval estimation. They define an interval within which the value of the estimated parameter is believed to be, with a stated percent confidence. The confidence level is usually high, at least 90%. The word "confidence" in this context is a very carefully defined technical term, the meaning of which is discussed below.

The sample mean is a particularly useful statistic and is used as the basis for the following discussion. If a sample of size n has been drawn on a random variable X, which has known variance σ^2, then a confidence interval at some specified confidence level α for the population mean μ may be estimated from

$$\bar{x} - w_\alpha \sigma/\sqrt{n} \leq \mu \leq \bar{x} + w_\alpha \sigma/\sqrt{n} \qquad (5\text{-}1)$$

If the population variance, σ^2, is unknown (which is the usual case) and must be estimated from the sample by the sample variance, s^2, (with some increase in average width), the confidence interval is:

$$\bar{x} - t_{\alpha,n-1}\, s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha,n-1}\, s/\sqrt{n} \qquad (5\text{-}2)$$

In the above, w_α and $t_{\alpha,n-1}$ are the values of a standardized variate having respectively the normal distribution (with n degrees of freedom) and the Student t-distribution (with n-1 degrees of freedom). As n increases, $t_{\alpha,n-1}$ approaches w_α. In both cases, they are the values of a variate symmetrically placed which enclose an area α between them as indicated below for the case of the normally distributed variate.

Interval in X $\mu - w_\alpha \sigma$ to $\mu + w_\alpha \sigma$	% Probability of Mean Being in Interval (α)
$\mu-\sigma$ to $\mu+\sigma$	68.27
$\mu-1.96\sigma$ to $\mu+1.96\sigma$	95.00
$\mu - 2\sigma$ to $\mu+ 2\sigma$	95.45
$\mu-2.57\sigma$ to $\mu+2.56\sigma$	99.00
$\mu-3\sigma$ to $\mu+3\sigma$	99.73

Andres (1984) discusses a procedure for establishing confidence intervals for a particular point (eg. the 95th percentile) on the cumulative distribution function (i.e. cumulative relative frequency). The method suggested by Andres is based upon the assignment of a binomial distribution to the height of the cumulative distribution function at each point of interest and then calculating a one-sided confidence interval based on this assignment. The underlying distribution does not need to be known, since the cumulative distribution function is estimated from the data by a non-parametric procedure.

The meaning of a confidence interval at confidence level α is a most important consideration. A confidence interval describes the unknown but constant value of a parameter in terms of an interval which is calculated from the observations of a sample. On repeatedly taking the same sized sample under the same conditions, and calculating confidence limits each time, the confidence limits will change from sample to sample while the true value of the parameter will remain constant. However, if the number of samples is indefinitely large, the fraction of confidence limits which enclose the true value approaches α. The actual stated confidence interval is always calculated from a single sample, but its explanation is in terms of the imagined scheme of indefinitely repeated sampling. Figures 5.1 to 5.4

show 95% confidence intervals about the mean.

An important property of the confidence intervals given by equations (5-1) and (5-2) is that, while their derivation depends on X being normally distributed, they are in fact very robust with respect to departure from normality. This means in practice that as long as the sample size is say 5 or more, little attention needs to be paid to the actual distribution of X. This property is a consequence of the Central Limit Theorem. (Recall for example the illustration of the Central Limit Theorem provided in Chapter 3.)

The range and various percentile estimates are other common measures of the dispersion in sample data. The range is simply the difference between the largest and smallest values in a sample. Percentile estimates can be obtained directly from the cumulative relative frequency histograms (or cumulative probability functions). For example, consider the lowest decile (10th percentile), the median (50th percentile), and the highest decile (90th percentile) marked on the cumulative relative frequency curve shown on Figure 5.5. These percentiles are points below which the output variable (Y5) was observed to lie 10%, 50% and 90% of the time respectively, in that sample. The difference between the 10th and the 90th percentile might be called the 80% probability limits and includes inherent uncertainty in the output variable (as well as residual sampling error from the relatively small number of trials - 500 in this case). This is different from the confidence interval on the mean discussed previously since the uncertainty described by the confidence interval results only from sampling error. It bears no relation to the inherent uncertainty in the variable which would be conveyed by its probability density function. For example, the confidence interval for the mean would be very narrow if a very large number of Monte Carlo trials had been used, even if the distribution of the random variable were broad.

5.3 Confidence Bands

Another type of confidence interval may be used to indicate the true cumulative probability function when the cumulative frequency diagram from a sample has been drawn. One of the most useful techniques for this is that of Kolmogorov and Smirnov (see for example Crow et al. 1960, or Conover 1980). It allows a staircase-like figure to be drawn above and below the cumulative frequency plot which establishes a confidence region for the cumulative probability function with basically the same meaning as the confidence intervals already described. That is, if the sampling were to be repeated indefinitely under the same conditions, the region between the confidence bands, which varies from sample to sample (decreasing with increasing sample size), would completely enclose the true cumulative probability function, which remains stationary, for a fraction of the samples which is numerically equal to the confidence level. The technique produces confidence bands of constant (vertical) width. The half-widths of these (vertical) confidence bands for selected quantities are shown below for the situation where 50 trials or more have been drawn.

Significance Level (α)	Half Width of Confidence Band
0.2	$1.07/\sqrt{n}$
0.1	$1.14/\sqrt{n}$
0.05	$1.36/\sqrt{n}$
0.01	$1.63/\sqrt{n}$

Figures 5.5 and 5.6 show the 95% confidence interval for the cumulative relative frequency and complementary relative frequency functions respectively. They are drawn as smooth curves although in actuality the graph is staircase-like with very small steps, 500 in this case. A weakness of this method is that it gives rise to bands which are of constant (vertical) width. This makes the band unnecessarily broad in the tails.

Techniques have been reported in the literature which indicate it is possible to reduce the width of the band in the tails. Loftis and Ward (1981) discuss how confidence intervals can be estimated when the variate (i.e. model output) is known to be normally distributed with a mean and variance estimated from the data. Unfortunately, the actual distribution of output values is not usually known. Cheng and Iles (1983) describe a method for establishing confidence bands for the probability density function. The technique first requires fitting a probability density function of known mathematical form to the observed relative frequency values by estimating the values of the parameters in the probability density function. Subsequently a joint confidence region is found for the (estimated) parameter values. The envelope of function relations obtained by substituting all parameter values within the joint confidence region into the probability density function is the confidence region (at the specified confidence level) for the probability density function of the variable. This technique has recently been extended to cumulative distributions and is undergoing further development (Reilly 1985).

A discussion of some related issues and their effect on the regulatory intepretation of the output of the probabilistic assessment is provided in the next chapter.

6.0 SOME ASPECTS OF REGULATORY CONCERN

6.1 Introduction

Previous chapters have discussed the sources of uncertainty in predictive modelling and various techniques to propagate these uncertainties through to the model outputs. This chapter touches on some aspects that might be of concern to regulators who will be required to assess the meaning and significance of these uncertainties.

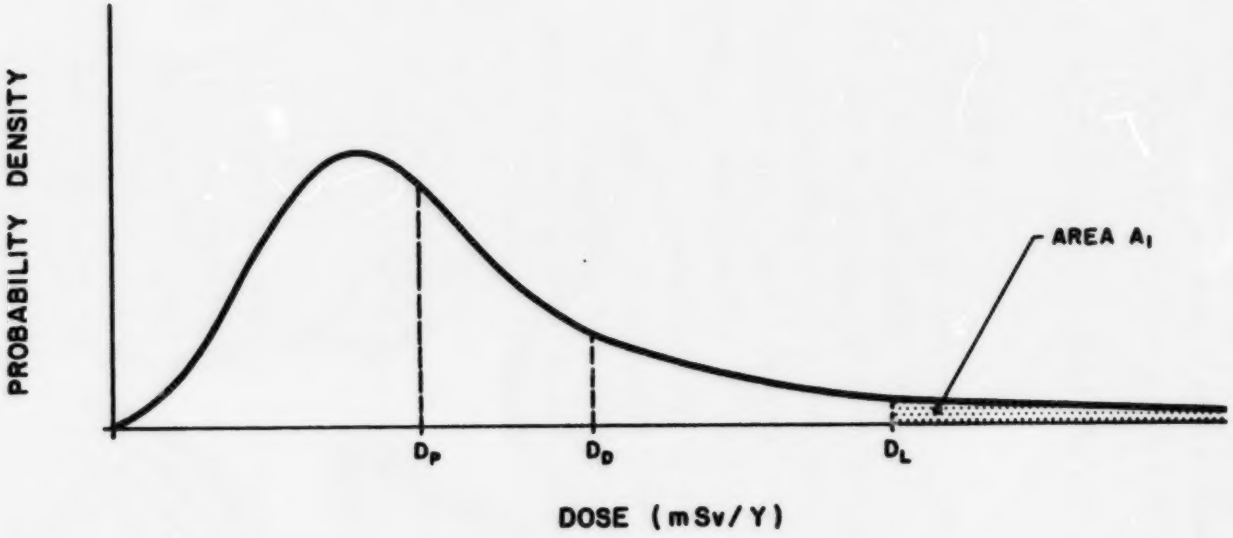
The interpretation of a probabilistic analysis, (that is an analysis which attempts to "accommodate" uncertainty and variability) is more complex than that of a deterministic analysis. Whereas a single level criterion, such as a dose limit, can be used to judge the acceptability of a consequence estimated by a (validated) deterministic analysis, in order to assess the results of a probabilistic analysis, decision makers and regulators must consider the uncertainties as well as the magnitude of the estimated consequences. In many instances, it is likely that a certain number of the probabilistic estimates will exceed the one-limit criterion that is used to judge the single estimate of consequence calculated by a deterministic analysis. Consequently, a more detailed criterion is required that explicitly acknowledges the existence of uncertainties associated with predictive modelling.

6.2 Interpretation of the Output of a Probabilistic Assessment

Prior to formulating any specific regulatory criteria, it is important that the output of a probabilistic assessment be properly interpreted. Consider for the moment Figure 6.1 which illustrates what might be a typical output of a probabilistic assessment. The output in the figure is shown as a smooth curve

FIGURE 6·1

TYPICAL OUTPUT
FROM A PROBABILISTIC MODEL



for this discussion although a probabilistic assessment actually produces a histogram as an output. Figure 6.1 is the curve which would result in the limit of an infinite number of runs of the probabilistic assessment model. Statements of confidence concerning a curve generated by a finite number of runs are discussed in Chapter 5. Figure 6.1 shows the annual dose to an average individual in the critical group, i.e. the group most likely to receive the maximum dose from the activity, facility, etc. being modelled. Assume for now that the characteristics that define the critical group (distance from facility, eating habits, etc.) are fixed; the issue of critical group will be discussed later.

The specific dose values noted on Figure 6.1 include: D_p, the expected or average dose estimated by a probabilistic analysis; D_d, the dose estimated by a deterministic (and probably conservative) analysis; and D_L, the annual dose limit for a member of the critical group - say 5 mSv/year for a member of the public. The area $(1-A_1)$ represents the degree of belief that one is entitled to hold that the actual dose will not exceed the dose limit D_L.

The interpretation of Figure 6.1 is quite dependent on the interpretation given to the distributions used to represent the uncertainties in the model input parameters. Assume for the moment that every input parameter is a constant and the input distributions represent our estimates of the degree of uncertainty about the true value of each input, i.e. each parameter has a "true" value but we are uncertain about what that true value is. The dose to the individual in the critical group is then also a fixed single value and the output of the probabilistic assessment represents our degree of uncertainty about that true value. Our "best guess" of the true value is the expected value or mean of the output distribution (i.e. D_P).

Furthermore, because of uncertainties, the probabilistic assessment indicates that the (subjective) probability that the annual dose to an average member of the critical group will not exceed D_L is $(1-A_1)$. For time-independent models (i.e. the scenario is fixed and does not change with time), this interpretation applies to all future years. It does not mean that the dose limit D_L will be exceeded in A_1% of future years - the true dose is fixed (because the inputs are fixed) and it does not mean that the dose will vary from year to year. The width of the output curve represents our uncertainty about that true value.

Consider a second case in which all the input parameters are truly random variables such as average wind speed or precipitation for any given year. The input distributions then represent the annual variability in each parameter (assume that this variability is well-known because we have many years of records on which to base our input distributions). The output distribution of the probabilistic assessment then also represents the year-to-year variability in the dose to an individual in the critical group. The dose varies from year to year because the inputs vary from year to year. Hence it can be said that in $(1-A_1)$% of all future years the dose will not exceed the dose limit.

In the first case noted above, the uncertainty in the predicted dose arises from our lack of knowledge of the constant values, in the other, from our inability to predict a variable. The uncertainty is there in both cases and is characterized by the output distributions. In both cases, the output distributions describe how strongly one is entitled to believe that an average member of the critical group will receive a particular dose in a particular year. In the first case, however, the dose will remain constant from year to year, in the other, it will vary.

Consider a third and probably more realistic case in which some input parameters are fixed but uncertain and some are truly variable (or indeed both variable and uncertain). The interpretation of the output distribution from a probabilistic assessment then falls somewhere between the two cases considered above.

The degree of confidence the modeller has about the value of A_1 (the probability of exceeding the dose limit D_L) might be of concern to a regulator. If it is assumed that the input distributions truly represent our estimate of the degree of uncertainty about the values of the input parameters, then our degree of confidence about the value of A_1 can be quantified using statistical methods, such as those discussed in Chapter 5. (These statistical methods are independent of the form of the model used in the probabilistic assessment. Our degree of confidence depends only on the statistics of the output distribution.) The value of A_1 becomes perfectly defined (in the sense that the input distributions are perfectly propagated through to the output distribution) as the number of model runs tends to infinity.

Of course, a "true" model does not exist since all models are only representations of reality. A regulator seeking to judge the acceptability of an operation characterized by the probabilistic assessment output shown in Figure 6.1 could be required to answer further questions. For example, a different group of experts could analyse the facility under study and carry out a probabilistic assessment that results in an output distribution different than Figure 6.1. This could arise because of their use of different input distributions for the model parameters or even the use of a different model. It may be determined by this second group of experts that the dose limit D_L is exceeded with a different probability A_2 (and different confidence intervals). Which probabilistic assessment does the regulator accept? Which value, A_1 or A_2, does the regulator accept?

These questions cannot be answered using quantitative (i.e. statistical) criteria. A quantitative confidence limit about the area A_1 that accommodates all model outputs cannot be formulated because no one can quantify the degree of precision that should be accorded different values of A_1 that are produced by different groups of experts (who use different models, etc.) It is up to the regulator to look at all the analyses - inputs, models, outputs - and use his best judgement, tempered by the credibility he places on each group of experts and their models, to accept or not accept the results. This is a subjective decision weighted by all the available information. No single regulatory quantitative criterion can be formulated to account for all the uncertainties because not all the uncertainties are knowable. Quantitative tests are available to determine if one output distribution is statistically different from another output distribution - this however does not tell the regulator which distribution is more representative of reality. In fact, the output distribution produced by each group of experts even if all are different are all correct by definition since each reflects the state of knowledge or level of confidence or degree of uncertainty that each group assigns to their output.

It can also be argued that to put too much credence into the "tails" of an output distribution (the region of low probability) or to define confidence limits about the value of A_1 is meaningless or at least unfruitful since input is always imperfectly known. Hoffman and Gardner (1984) argue that "meaningful estimates of maximum or extreme values in the tails of predicted distributions are currently beyond the capabilities of extant models and data sets." They find it most productive to focus attention on a "probable range" of predicted values, since this is a range that is reasonably likely to encompass a "true" value. They would go as far as to claim that, given the nature of the models and input parameter distributions, the extreme values have no physical significance and do not represent "real cases".

Some investigatiors suggest that it is not useful to develop strict quantitative criteria that focus undue attention on our degree of confidence about the tails of the distribution. It might be better to develop a less specific criterion that requires a probabilistic assessment analysis to demonstrate that there is a "high probability" that the dose to the individual in the critical group will not exceed a specified limit D_L.

There are others who argue that it is proper and desirable to examine the tails of the output distribution. They contend that the tails of the input distributions represent our lack of knowledge about the input parameters. Consequently, the tails of the output distribution, the area A1 in Figure 6.1, represents our degree of belief that a particular dose limit will be exceeded. Under this interpretation, it might then be valid to formulate a criterion such that the probability (our degree of belief) that the dose limit D_L will be exceeded will not be greater than a specified value.

6.3 The Concept of the Critical Group

The variability about the characteristics that define the critical group was alluded to earlier. This variability is not specifically accounted for in present regulations other than in the definition of critical group which requires the critical group to be chosen so that the doses expected to be received by the group (as calculated by a deterministic analysis) are higher than the average received by typical people in all other groups. Under present guidelines, this would preclude using variable input characteristics for a critical group in a probabilistic analysis that would result in a large range of expected doses, e.g. allowing the distance between the group and the source of exposure to be highly variable. This approach, in effect, would re-define the critical group with each distance and the whole concept of critical group would be lost.

7.0 CONCLUDING REMARKS

There are many factors which affect the accuracy and precision of quantities calculated using probabilistic methods. This report discussed many of these sources of uncertainty and various ways of addressing them. In particular, the report examined various ways in which uncertainty in model input parameters can be propagated through a model to provide a distribution of output values. Except for cases where models can be reduced to very simple equations, analytic propagation of uncertainty is impractical. In our view, the most appropriate approach, at least for environmental models, is to use Monte Carlo techniques, specifically techniques which employ either simple random sampling or latin hypercube sampling strategies.

Many tools are available to examine how sensitive of uncertain model outputs are to changes or uncertainties in model inputs including graphical methods, various sensitivity and uncertainty indices and regression techniques. Our investigations suggest that no single technique is universally applicable. However, regression analysis is perhaps the single most powerful technique and provides insight as to both how sensitive and uncertain model predictions are to specific parameters. Notwithstanding this observation, all of the available tools provide insight to model interpretation and as many techniques as possible should be applied as permitted within the constraints of a particular study.

The degree of confidence that can be placed in probabilistic predictions is an important consideration. Techniques are available for providing estimates of confidence concerning either a single output descriptor (eg. mean) or the entire output distribution.

Regulatory evaluation of the predictions of a probabilistic assessment is complex. The regulator is faced with considering an entire curve of output values, rather than a single value. Moreover, science alone cannot help the regulator to assess the "correctness" of a particular probabilistc assessment. Expert judgement is also required.

REFERENCES

Andres, T.H., 1985. Personal Communication.

Andres, T.H., 1984. "Confidence Bounds on an Empirical Distribution." Atomic Energy of Canada Limited, Draft. (Feb. 1984).

Beckmann, P., 1968. "Elements of Applied Probability Theory." Harcourt, Brace and World, Inc., New York.

Boehm, B. W., 1984. "Verifying and Validating Software Requirements and Design Specifications." IEEE Software, Vol. 1, No. 1.

Cameron, F.X., 1982. "The Problem of Human Intrusion; Towards a Resolution of Uncertainty." In Proceedings of the Symposium on Uncertainties Associated with the Regulation of the Geologic Disposal of High Level Radioactive Waste. Gatlinburg, Tennessee, March 9-13, 1981. NUREG/CP-0022.

Cheng, R.C.H. and Iles, T.C., 1983. "Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables." Technometrics, Vol. 25, No. 1, pp. 77-86.

Cochrane, W.G., 1977. "Sampling Techniques." John Wiley and Sons, New York.

Collee, R., Abee, H.H., Cohen, L.K., Ed, D., Eisenhower, E.H., Jarvis, A.N., Fisenne, I.M., Jackson, M., Johnson, R.H., Jr., Olson, D., and Peel, J., 1980. "Reporting Environmental Radiation Measurement Data." pp. 6-1 through 6-34 in "Upgrading Environmental Radiation Data", Health Physics Society Committee Report HPSR-1, EPA 520/1-80-012, U.S. Environ. Protection Agency.

Conover, W.J., 1980. "Practical Non-Parametric Statistics." (second edition), John Wiley and Sons.

Cranwell, R.M. and Helton, J.C., 1981. "Uncertainty Analysis for Geologic Disposal of High Level Radioactive Waste." SAND-81-1718C, Sandia National Laboratories, Albuquerque, N.M.

Crow, E.L., Davis, F.A. and Maxfield, M.W., 1960. "Statistics Manual." Dover Press.

Downing, D.J., Gardner, R.H., and Hoffman, F.O., 1985. "An Examination of Response-Surface Methodologies for Uncertainty Analysis in Assessment Models." Technometrics, Vol. 27, No. 2, pp. 151-163, May.

Draper, V. and Smith, H., 1981. "Applied Regression Analysis." Second Edition, John Wiley & Sons Inc., New York.

Edwards, A.L., 1979. "Multiple Regression and the Analysis of Variance and Covariance." W.H. Freeman & Co.

Feller, W., 1968. "An Introduction to Probability Theory and Its Applications." Vol. 1, Third Edition, John Wiley and Sons, Inc., New York.

Fox, D.F., 1984. "Uncertainty in Air Quality Modeling." A Summary of the AMS Workshop on Quantifying and Communicating Model Uncertainty, Woods Hole, Mass, Sept. 1982. Bulletin American Meteorological Society, Vol. 65, No. 1, January, pp. 27-36.

Gardner, R.H., O'Neill, R.V., Mankin, J.B., and Carney, J.H., 1981. "A Comparison of Sensitivity Analysis and Error Analysis Based on a Stream Ecosystem Model." Ecological Modelling, Vol. 12, pp. 177-194.

Gardner, R.H., O'Neill, R.V. and Hoffman, F.O., 1982. "Assessing Model Uncertainties." In Proceedings of the Symposium on Uncertainties Associated with the Regulation of the Geologic Disposal of High Level Radioactive Waste. Gatlinburg, Tennessee, March 9-13, 1981. NUREG/CP-0022.

Gardner, R.H., 1984. "A Unified Approach to Sensitivity and Uncertainty Analysis." Proceedings of the 10th ISTED (The International Association of Science and Technology for Development), International Conference, Applied Simulation and Modelling, San Francisco.

Gardner, R.H., 1985. (in press). "Error Analysis and Sensitivity Analysis." Encyclopedia of Systems and Control, M. Singh (ed.), Pergamon Press, London.

Gentle, J.E., 1981. "Portability Considerations for Random Number Generators." IMSC Technical Report Series, No. 8101, February.

Helton, J.C., and Iman, R.L., 1982. "Sensitivity Analysis of a Model for the Environmental Movement of Radionuclides." Health Physics, Vol. 42, No. 5, pp. 565-584.

Hendrickson, R.G., 1984. "A Survey of Sensitivity Analysis Methodology." U.S. Department of Commerce, National Bureau of Standards, Report NBSIR 84-2814, Washington, DC.

Henrion, M. and Morgan, M.G., 1984. "A Survey of Techniques for Uncertainty Analysis." Presented at Workshop on Uncertainty Analysis, Carnegie-Mellon University, Pittsburgh, October 11-12.

Hofer, E., and Krzykacz, B. G., 1984. "Modelling and Propagation of Uncertainties in the German Risk Study." in Hofer E., "Probabilistic Uncertainty Analysis of PRA"ISPRA Course on Uncertainty Modelling in Probabilistic Risk Assessment." Ispra, November 1984.

Hoffman, F.O. and Gardner, R.H., 1983. "Evaluation of Uncertainties in Environmental Radiological Assessment Environmental Dose Assessment Models." In: Radiological Assessment: A Textbook on Environmental Dose Assessment, J.E. Till and H.R. Meyer (eds.), NUREG/CR-3332, ORNL-5968. U.S. Nuclear Regulatory Commission, Washington, D.C.

Hoffman, F.O. and Miller, C.W., 1983. "Uncertainties in Environmental Radiological Assessment Models and Their Implications." In Proceedings of the Nineteenth Annual Meeting of the National Council on Radiation Protection and Measurement (NCRP), Washington, DC.

Hoffman, F.O. and Gardner, R.H., 1984. "Parameter Uncertainty Analysis Application - A Position Paper." Prepared for SENES Consultants Limited under Supply and Services Canada Contract No. OSQ83-00236.

Iman, R.L. and Conover, W.J., 1979. "The Use of the Rank Transform in Regression." Technometrics, Vol. 21, No. 4, pp. 409-599, November.

Iman, R.L., and Conover, W.J., 1980. "Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment." Communications in Statistics: Theory and Methods, Vol. A9, No. 17, pp. 1749-1842.

Iman, R.L., Davenport, J.M., and Zeigler, P.K., 1980. "Latin Hypercube Sampling (Program Users Guide)." SAND79-1473, Sandia National Laboratories, Albuquerque, New Mexico.

INTERA Environmental Consultants Inc., 1983. "A Proposed Approach to Uncertainty Analysis." Prepared for Office of Nuclear Waste Isolation (ONWI), Battelle Memorial Institute, Columbus, Ohio, ONWI-488.

Kaplan, S., 1981. "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations - Application to Seismic Risk Assessment." Risk Analysis, Vol. 1, No. 3, pp. 189-196.

Kaplan, S. and Garrick, J. B., 1981. "On the Quantitative Definition of Risk." Risk Analysis, Vol. 1, No. 1, pp. 11-27.

Kleijnen, J.P.C., 1974. "Statistical Techniques in Simulation - Part 1." Marcel Dekker, Inc., New York.

Ku, H.H., 1966. "Notes on the Use of Propagation of Error Formulas." J. of Res. of the NBS - C. Engineering and Instrumentation, Vol. 70C, No. 4, October-December, pp. 331-341.

Little, A.D., Inc., 1983. "Research Needs for Dealing with Uncertainty in Risk Analysis." NTIS No. PB84-150127. Prepared for National Science Foundation Washington, DC.

Loftis, J.C. and Ward, R.C., 1981. "Evaluating Stream Standard Violations Using a Water Quality Data Base." Water Resources Bulletin, Vol. 17, No. 6, p. 1071-1077.

McKay, M.D., Beckman, R.J. and Conover, W.J., 1979. " A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." Technometrics, Vol. 21, No. 2, pp. 239-245.

Morse, K. and Roberts, S.D., 1983. "Implementing a Portable FORTRAN Uniform (0,1) Generator." Simulation, pp.135-139, October

Myers, G. J., 1979. "The Art of Software Testing." Wiley (NY).

Oblow, E.M., 1983. "An Automated Procedure for Sensitivity Analysis Using Computer Calculus." Oak Ridge National Laboratory, ORNL/TM-8776.

O'Neill, R.V., 1971. "Error Analysis of Ecological Models." Oak Ridge Nat. Lab., Radionuclides in Ecosystems. Proceedings of the Third Nat. Symp. on Radioecology, May.

Raj, D., 1968. "Sampling Theory." McGraw Hill, New York.

Reilly, P.M., 1985. "Discussion Paper on Confidence Regions for Output Distributions from Monte Carlo Calculations." Prepared for SENES Consultants Limited under Supply and Services Canada Contract 0SQ84-00207.

Rose, K.A., and Swartzman, G.L., 1981. "A Review of Parameter Sensitivity Methods Applicable to Ecosystem Models." NUREG/CR-2016, U.S. Nuclear Regulatory Commission, Washington, DC.

Rubinstein, R.Y., 1981. "Simulation and the Monte Carlo Method." John Wiley and Sons, New York.

Satterthwaite, F.E., 1959. "Random Balance Experimentation," Technometrics, Vol. 1, No. 2, pp. 111-137.

Seiler, F.A., 1984. "Analytical Methods: Advantages and Limitations." Presented at Workshop on Uncertainty Analysis, Carnegie-Mellon University, Pittsburgh, October 11-12.

SENES Consultants Limited, 1984. "A Study to Develop and Demonstrate Techniques for the Probabilistic Assessment of the Long-Term Effects of Uranium Mill Tailings in Canada - Phase I." Research Report prepared for the National Uranium Tailings Program, CANMET, EMR, under Supply and Services Canada Contract No. 0SQ83-00236.

SENES Consultants Limited, 1985. "Approaches to Risk Assessment for Canadian Uranium Mill Tailings." Report to the National Uranium Tailings Program, Supply and Services Canada Contract No. 0SQ84-00335 (In Preparation).

Smith, C.F., and Cohen, J.J., 1982. "Models and Criteria for Waste Repository Performance." In: Proceedings of the Symposium on Uncertainties Associated with the Regulation of the Geologic Disposal of High Level Radioactive Waste. Gatlinburg, Tennessee, March 9-13, 1981. NUREG/CP-0022.

Taylor, J.R., 1982. "An Introduction to Error Analysis." University Science Books, Mill Valley, California.

Tomovic, R. and Vukobratovic, M., 1963. "General Sensitivity Theory." American Elsevier.

Wuschke, D.M., Rice, A.M., and Gillespie, P.A., 1983. "Environmental Assessment of Sub-seabed Disposal of Nuclear Wastes: A Demonstration of Probabilistic Systems Analysis." Atomic Energy of Canada Ltd., Pinawa, Manitoba, AECL TR-206.

APPENDIX A: NOTES ON SAMPLING FROM RANDOM DISTRIBUTIONS

When carrying out probabilistic modelling, one is required to obtain random samples from a large range of statistical distributions. For example, one may wish to sample the moisture content of a tailings cover where it has been assumed that the moisture content is normally distributed. The following describes by way of example a general method, sometimes called the inverse transform method (or the probability integral transformation), that can be used to carry out such random sampling. Detailed descriptions of this and other methods, such as the composition method or the acceptance-rejection method, can be found in several texts, including Rubinstein (1981).

The basis of all methods of random sampling is the ability to draw random samples from a uniform distribution U(0,1) i.e. have equal probability of choosing any number between 0 and 1. In what follows, it is assumed that this capability is available. (Computers generally have a built-in uniform random number generator.)

The inverse transform method makes use of the cumulative distribution function F(x) of a random variable x and is defined by

$$F(x) = \int_{-\infty}^{x} f(x')\,dx' \qquad \text{(A-1)}$$

where f(x) is the probability density function for the random variable, x. Thus, F(x) is just the cumulative probability that the random number chosen from the probability distribution f(x) is less than or equal to x. (Some statisticians prefer the simpler term "distribution function" for F(x) since they consider the adjective "cumulative" redundant (e.g. Feller 1968).

For example, Figure A.1 shows the exponential probability density function

$$f(x) = ae^{-ax} \qquad (x \geq 0) \qquad (A-2)$$

and its cumulative distribution function

$$F(x) = \int_{0}^{x} ae^{-ax'}\, dx' = 1 - e^{-ax} \qquad (x \geq 0) \qquad (A-3)$$

Note that f(x) meets the necessary conditions for a probability density function: f(x) is non-negative for all x and f(x)dx integrates to unity.

It can be seen that while the range of f(x) is determined by the particular form of f(x), the range of the cumulative distribution function F(x) is always 0 to 1. This is a consequence of the definition of F(x) and holds for all probability density functions (i.e. all f(x)'s).

This transformation of ranges enables the sampling of a random number from f(x) as follows. First, a number from the uniform distribution U(0,1) is chosen (say 0.4). Graphically, this number is found on the vertical axis of the cumulative distribution (curve F(x), Figure A.1b). The number selected from the vertical axis (0.4 in this example) has a corresponding value on the horizontal axis, namely x = 0.26, and this value is the random number drawn from possible values of x in accordance with the exponential probability density function f(x).

Algebraically, to obtain a random value x from the exponential distribution, we choose a value u from the uniform distribution U(0,1) and solve

FIGURE A·1a

EXPONENTIAL
PROBABILITY DENSITY FUNCTION, f(x)

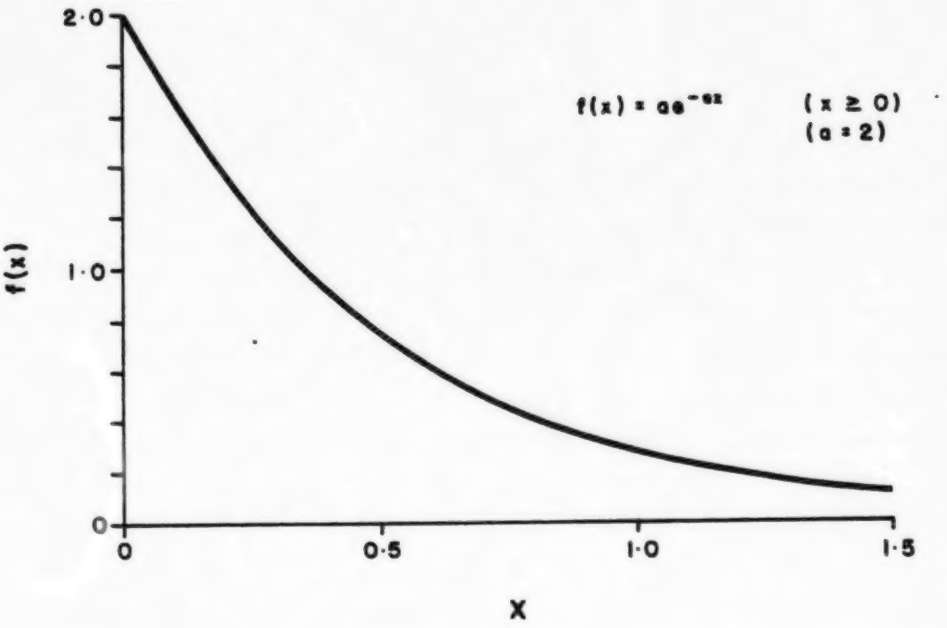
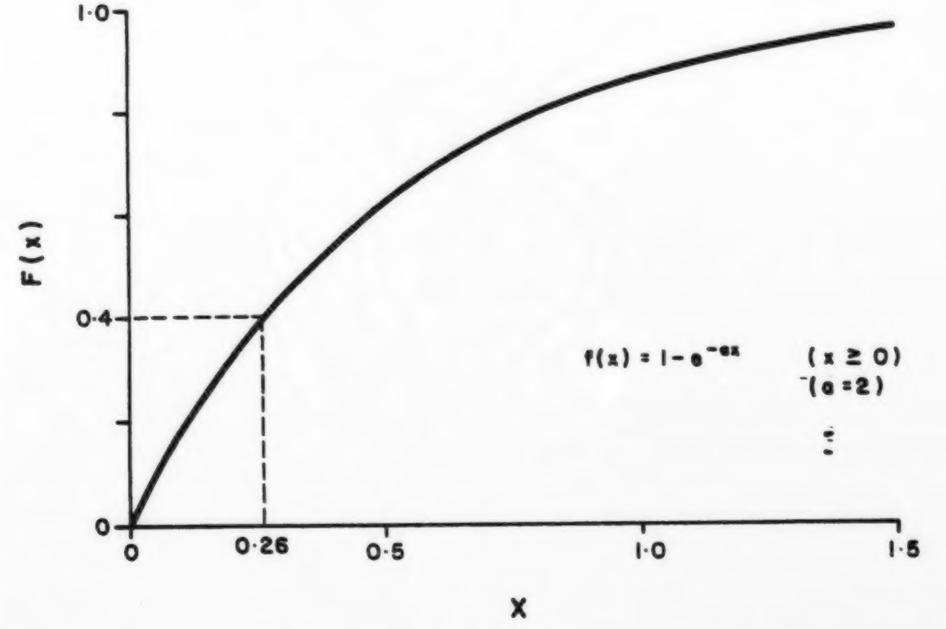


FIGURE A·1b

CUMULATIVE
DISTRIBUTION FUNCTION, F(x)



$$F(x) = u = 1-e^{-ax} \qquad (A-4)$$

or

$$x = -\ln(1-u)/a \qquad (A-5)$$

That is, if u is chosen from the uniform distribution U(0,1), the corresponding x drawn from possible values in accordance with the exponential distribution, $f(x) = ae^{-ax}$, is given by equation (A-5). Since the value of x determined by equation (A-5) results from inverting equation (A-4), this technique of drawing a sample is sometimes called the inverse transform method. (It is simple to show (Kleijnen 1974) that (1-u) has the same probability density distribution as u and thus equation (A-5) can be replaced by $x = -\ln(u)/a$. This is useful if computer time is important since one less subtraction is required.)

Consider a second example in which we wish to generate a random sample from a different uniform distribution, U(a,b). In this case,

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \qquad (A-6)$$

and

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \qquad (A-7)$$

We generate u from U(0,1) and solve

$$F(x) = u$$

or $$x = a + (b-a)u \tag{A-8}$$

and x is the required sample from U(a,b). Note that in the trivial case of a=0 and b=1, we have x=u as expected.

As noted earlier, this is a general method that can be used for any probability density function f(x). However, it is not always possible to analytically determine the cumulative distribution function F(x) by integrating f(x) and then inverting the cumulative distribution function for the selected uniformly distributed value, as done in the above example. (Probability density functions that can be solved analytically include exponential, uniform, Weibull, logistic and Cauchy.) Although numerical integration and inversion may be used, methods based on the particular properties of the statistical distribution chosen are often more convenient and/or more efficient. One such method for the very important normal distribution is discussed below. (Methods specific to other functions such as binomial, poisson and beta can be found in Rubinstein (1981).)

The method for generating random numbers from a normal distribution is an approximate method and is based on the Central Limit Theorem of Statistics. In its simple form, the Central Limit Theorem states that the distribution of the sum of several independent variables (regardless of the type of distributions from which they are drawn)

$$x = x_1+x_2+...+x_n \tag{A-9}$$

will asymptotically approach the normal distribution as the number of variables increases (i.e. as n gets large). Moreover, the mean and variance of the distribution of x is the sum of the drawn. A necessary condition for the theorem to hold is that the variance of any one distribution is negligible compared to the sum of the variances of all the distributions.

To demonstrate the application of the Central Limit Theorem to the generation of a variable from a normal distribution, let's first consider sampling a variable from a uniform distribution. It can be shown easily that the uniform distribution U(0,1) has a mean equal to 1/2 and a variance equal to 1/12. Thus, the sum of n independent uniform variables has mean n/2 and variance n/12. By the Central Limit Theorem, the sum is approximately normally distributed if n is sufficiently large. For almost all purposes, n = 12 is adequate (Rubinstein 1981) and the sum of 12 independent uniform variables, each from U(0,1), is approximately normally distributed with mean 6 and variance 1 using this method. By subtracting 6 from the sum, the mean is 0 and the variance is 1. Using this method, one sample of the variable x would require 12 samples from U(0,1).

If random samples are required from a normally distributed variable z which has mean $\bar{z}$ and variance σ_z^2, one can derive z from x (which has mean 0 and variance 1) by the relation

$$z = \bar{z} + x\sigma_z \tag{A-10}$$

Thus, a random variable can be chosen from any normal distribution one wishes.

This method can be extended easily to sample from a lognormal distribution with geometric mean μ_g and geometric standard deviation σ_g. Since the logarithms of variables from a lognormal distribution are normally distributed, one first samples from a normal distribution (as shown above) with mean $\ln(\mu_g)$ and standard deviation $\ln(\sigma_g)$. The sample x is then transformed back to logarithm space by e^x; that is, e^x is the required lognormally distributed random variable.

APPENDIX B: THE METHOD OF DISCRETE PROBABILITY DISTRIBUTIONS

B.1 Introduction

The method of Discrete Probability Distributions (DPD's) is another approach to the problem of combining or propagating input distributions through a mathematical model to obtain distributions for the desired output quantities. The DPD method is an alternative to the random sampling or Monte Carlo methods described in Section 3.4. In spirit, the DPD method has some of the flavour of the systematic sampling method of Section 3.4.5 and of the so called "importance weighted" Monte Carlo methods. The basic idea is to reduce the amount of labour required of the computer by injecting into the procedure some knowledge of the structure of the particular problem being solved. The basic technique and philosophy of the DPD method are described in Kaplan (1981).

Proponents of DPD suggest that the advantages of the approach are that it is a finite calculation, it is not open ended, it is not subject to sampling error, and it may require less computer time than Monte Carlo. They also argue that it allows one to concentrate the computational effort in the area of interest, such as the upper tails. One disadvantage is that it requires more preparation of the discretized input.

The proponents of the DPD procedure also argue that the technique is extremely flexible in that it is able to encompass uncertainties of very general types. (This is also true of Monte Carlo techniques.) That is, in DPD the entities about which we are uncertain need not be ordinary simple scalar variables. They can be more general entities, for example, vectors, matrices, curves, functions, entire models, and sets of equations describing physical phenomena. One example of this is the treatment of seismicity and fragility curves in the seismic risk assessment described by Kaplan (1981).

The remainder of this section gives the results of applying the DPD method to the five examples discussed in Section 3.4. The results are compared to those derived with simple random sampling.

B.2 Examples

The following assumes the relationships between input variables A,B,C,D,E and output variables Y1 to Y5 as summarized in Table 3.2.

Consider now that we are uncertain about the values of the input variables, and have described our state of uncertainty by assigning to each a probability distribution as specified in Table 3.3. What state of knowledge about the output variables is consistent with, or dictated by, this uncertainty in the input variables? Stated another way, we seek to "propagate" the distributions of Table 3.3 through the model equations Y1 to Y5.

The first step is to discretize the input variables A,B,C,D,E. To proceed, first note that the lognormal variable D may be written in terms of the standard normal variate as follows:

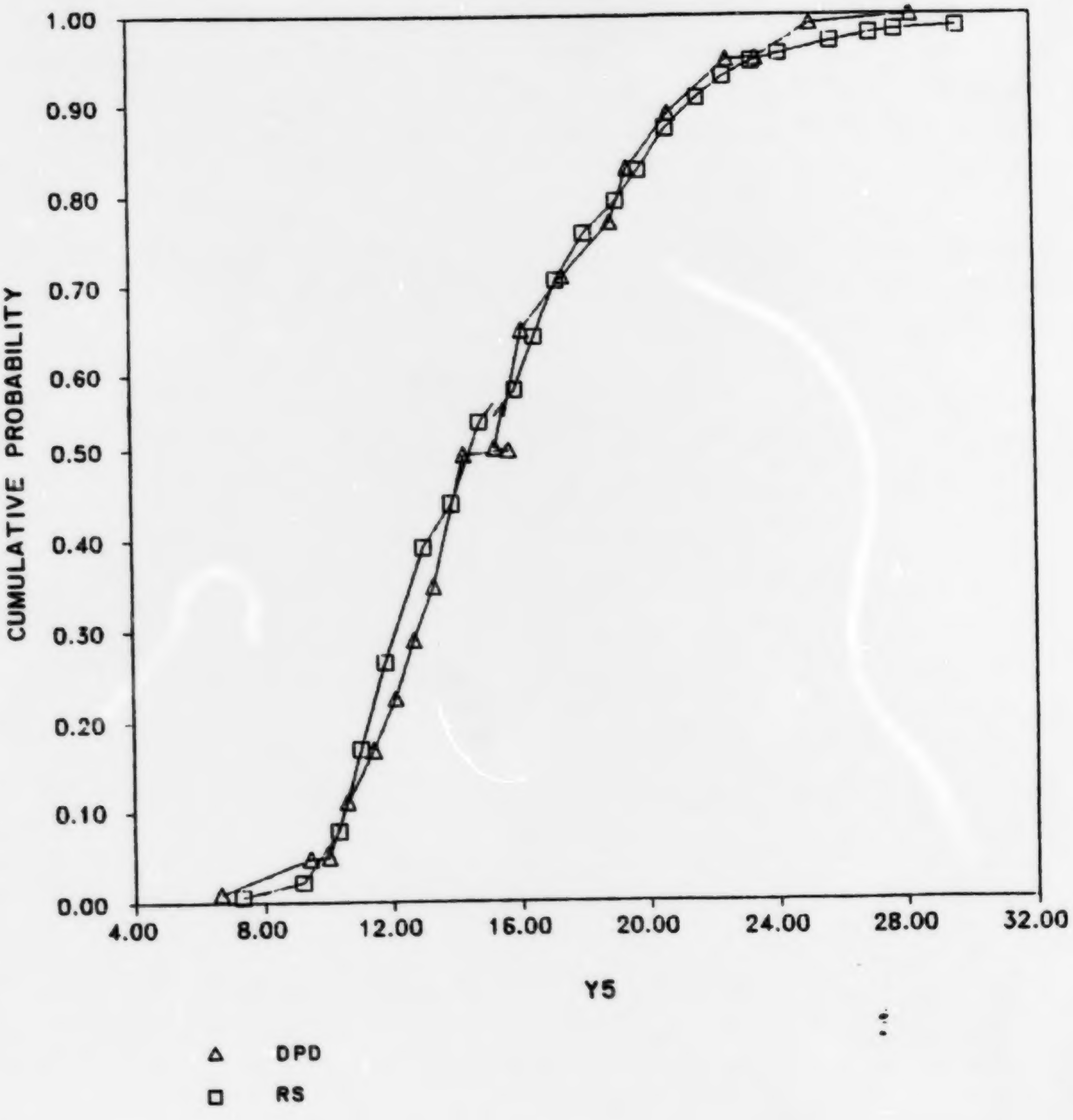
$$D = 5.66\ (1.2)^{Z} = 5.66\ e^{0.182Z}$$

Similarly, the normal variable E may be written in terms of a second standard normal variate S as follows:

$$E = 1.5 + 0.5^{\ S}$$

FIGURE B·1

CUMULATIVE HISTOGRAMS OF MODEL Y5 CALCULATED USING RS AND DPD



Each of these input variables were discretized into 20 point DPD's.

The equations for Y1 to Y5 can be rewritten in terms of the five input variables A,B,C,Z and S (Z and S have of course the same distribution). Using the technique described by Kaplan (1981), the discretized input variables were propagated through equations Y1 to Y5. The mean and variance of the output distributions are shown in Table B.1. Also shown in Table B.1 are the corresponding means and variances calculated using simple random sampling.

For purposes of illustration, the cumulative distributions for realizations of equation Y5 are shown in Figure B.1 for both DPD and simple random sampling.

Table B.1

Comparison of Means and Variances of Example Models Calculated Using DPD and RS

	DPD[1]		RS[2]	
Model	Mean	Variance	Mean	Variance
Y1	24.1	3.09	24.2	3.33
Y2	786	1.71×10^5	824	1.91×10^5
Y3	65.7	111	66.3	120
Y4	4590	4.69×10^7	5430	6.81×10^7
Y5	15.8	18.1	15.7	19.2

Notes

1. Based on 500 trails using the method of discrete probability distributions (DPD).
2. Based on 500 trials using simple random sampling (RS).